[illegible]YCHOLOGICAL REVIEW[illegible] [illegible]HOLE NO. [illegible]
[illegible]1933

# [illegible]ological [illegible]raphs



# [illegible]perimental S[illegible] [illegible]actors [illegible]cing Conson[illegible] [illegible]ments

BY

EUGENE GOW[illegible]

[illegible]ROM THE PSYCHO[illegible] [illegible]BORATORY
OF VANDERBILT [illegible]

[illegible] FOR THE AMERICAN P[illegible] [illegible]OCIATION BY
[illegible]YCHOLOGICAL RE[illegible] [illegible]ANY
PRINCETON[illegible]
AND ALBAN[illegible]

# P[illegible] of the
# American [illegible]cal Association

EDWARD C[illegible] UNIVERSITY (*Review*)
[illegible] W. FERNBERGER [illegible]SYLVANIA (*J. of Exp. Psych.*)
WALTER S. HU[illegible]ITY (*Index and Abstracts*)
HENRY T. MOO[illegible] (*J. Abn. and Soc. Psychol.*)
HERBERT S. LA[illegible] UNIVERSITY (*Monographs*)
EDWARD [illegible] UNIVERSITY (*Bulletin*)

HERB[illegible] Business Editor

[illegible]VIEW

[Co]ntaining original co[illegible] bi-monthly, January, March, May, July, September, and [illegible]rs comprising a volume of about [illegible] pages.

[illegible]LLETIN

[Co]ntaining critical rev[illegible]s, psychological news and notes, university notices, and [illegible] monthly (10 numbers), the annual volume comprising ab[illegible] issues of the BULLETIN consist of general reviews of [illegible] ment of psychology.

JOUR[illegible]AL PSYCHOLOGY

[Co]ntaining original cont[illegible] of [illegible]ntal character, appears bi-monthly, February, April, June, A[illegible] O[illegible] December, the six numbers comprising a volume of ab[illegible] pages [illegible] 1, 1933).

[illegible] INDEX

[Is] a compendious bib[illegible]graphs, and articles upon psychological and cognate topic[illegible]d during the year. The INDEX is issued annually in [illegible]ribed for in connection with the periodicals above, or [illegible]

[illegible]STRACTS

[Ap]pears monthly, the [illegible]ndex supplement making a volume [of] about 700 pages. [illegible] to the publication of non-critical abstracts of the world[illegible]gy and closely related subjects.

[illegible]NOGRAPHS

[Co]nsist of longer rese[illegible]ctions of laboratory studies which it is important to pu[illegible]nits. The price of single numbers varies according to their [illegible]APHS appear at irregular intervals and are gathered into vol[illegible] pages.

JOURNAL OF [illegible] SOCIAL PSYCHOLOGY

[Ap]pears quarterly, [illegible] March, the four numbers comprising a volume of [illegible] contains original contributions in the field of abnormal and [illegible]s, notes and news.

[illegible]ON RATES

[R]eview: $5.50 (Foreig[illegible] per volume.
Journal: $7.00 ([illegible]: $6.00 per volume (Foreign, [illegible]),
Bulletin: $[illegible]sts: $8.00 (Foreign, $8.25).
Abnormal [illegible], $5.25). Single copies $1.50.
Current numbers[illegible]$1.00; Abstracts, 75c; Bulletin, 60c.

[illegible]RATES

Review and [illegible], $10.50).
Review and [illegible], $11.50).
Bulletin and [illegible], $12.50).
Review, Bu[illegible] (Foreign, $16.75).
Review, Bu[illegible]: $19.00 (Foreign, $19.75).

Subscriptions, o[illegible]nications should be sent to the

## PSYCH[illegible]VIEW COMPANY

[illegible]N. J.

VOL. XLV No. 2 PSYCHOLOGICAL REVIEW PUBLICATIONS WHOLE NO. 201 1933

# Psychological Monographs



# An Experimental Study of Factors Influencing Consonance Judgments

BY

EUGENE GOWER BUGG

A STUDY FROM THE PSYCHOLOGICAL LABORATORY OF VANDERBILT UNIVERSITY

Published for The American Psychological Association by

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, N. J.
and ALBANY, N. Y.

## ACKNOWLEDGMENTS

This study was carried out in the psychological laboratory of Vanderbilt University, under the direction of Dr. Lyle H. Lanier. The writer wishes to thank Dr. Lanier for his untiring criticism, both during the course of the experimental work and during the preparation of this report. Acknowledgment is also made of obligation to Dr. F. C. Paschal, professor of psychology of Vanderbilt University, for various suggestions concerning the style of presentation of the present study; to Dr. Browne Martin of the Nashville Conservatory of Music for the loan of various books; to my wife Ophelia Ragsdale Bugg for assistance in the preparation of the manuscript; and to my friend and former vocal instructor Signor G. S. de Luca, President of the Nashville Conservatory of Music.

The writer desires to acknowledge his great indebtedness to Dr. Herbert C. Sanborn, head of the department of philosophy and psychology of Vanderbilt University, both for his inspiring instruction in philosophy, especially as regards the general methodology of science, and for many valuable suggestions concerning the present study.

CONTENTS

PAGE

Chapter I. INTRODUCTION ........................... 1

Chapter II. HISTORICAL SURVEY ........................ 8
1. Theories of consonance................ 9
2. Experimental studies .................. 19

Chapter III. PROBLEMS AND EXPERIMENTAL METHODS OF THE PRESENT STUDY................. 28
1. The Seashore Consonance Test, with "consonance" directions ............. 30
2. The Seashore Consonance Test, with "preference" directions ............. 32
3. Four "consonance tests" of varying difficulty .............................. 33
4. The preferential use of three criteria..... 37
5. The use of each of two criteria in separate series ............................ 39

Chapter IV. GENERAL ANALYSIS OF THE SEASHORE CONSONANCE TEST "SCORES"........... 42
1. The general efficiency of the subjects on the "consonance" and "preference" tests 42
2. Comparative efficiencies of musically trained and untrained subjects................ 44
3. The general reliability of the "consonance" and "preference" tests........ 45
4. Comparative reliabilities of musically trained and untrained subjects......... 47

Chapter V. THE EFFECTS OF PAIRED-INTERVAL DIFFICULTY ON CONSONANCE JUDGMENTS.......... 49
1. The relative difficulty of the combinations of intervals in the Seashore Consonance Test ............................ 51
2. Analytical study of the influence of difficult pairs of intervals on reliability..... 56
3. The reliability of consonance judgments in separate tests of varying difficulty...... 62

PAGE

Chapter VI. THE INFLUENCE OF AFFECTIVE-TONE UPON CONSONANCE JUDGMENTS ............ 68
1. General comparisons of consonance and preference judgments ............... 71
2. An analysis of the effect of order of presentation of intervals upon "consonance" and "preference" judgments.......... 72

Chapter VII. THE INFLUENCE OF DIFFERENT CRITERIA ON THE CONSISTENCY OF CONSONANCE JUDGMENTS ...................... 79
1. The "preferential" use of three criteria.. 82
2. The use of each of two criteria in separate series ........................... 84

Chapter VIII. GENERAL SUMMARY ..................... 90

Chapter IX. CONCLUSIONS ........................ 96

REFERENCES............................................ 99

# CHAPTER I

## Introduction

Few problems in audition have provoked so much controversy and so little experimental research as the problem of "consonance." Psychological literature on this subject has been limited for the most part to "theories" of consonance perception, the majority of them deriving almost entirely from *a priori* speculation. As early as 600 B.C. Pythagoras observed that the relations of "consonant" tones could be represented by small whole numbers. That this observation uncovered an important problem for psychology no one can deny. But that the problem itself could have been accepted in large measure as its own solution by many thinkers in this field seems scarcely credible. Yet this is substantially what occurred in the views of Leibniz, Euler, Schopenhauer and—to a considerable degree—Lipps. These men lived, however, in the days of pre-scientific psychological explanation and perhaps should not be held too strictly to account for explaining consonance in terms of processes or entities which merely symbolized in this or that form the fact observed by Pythagoras. With the development of a scientific sense-physiology and psychology in the nineteenth century it was inevitable that attention should be turned to such an important auditory relationship as the "consonance" of tones. Helmholtz was probably the initiator of the scientific approach to the problem in his attempt to relate the phenomena observed both to physical stimuli and to physiological processes; but he probably over-simplified both the problem and the explanation—as both his rationalistic predecessors and also most of his successors have done. However, he brought to the problem an empirical point of view which regarded it as one susceptible of experimental attack. The fact that so little fundamental insight into this problem has been developed subsequent to Helmholtz can probably be attributed to the failure to be guided by his scientific precepts.

An ever increasing number of experimental studies bearing on consonance perception have appeared since the time of Helmholtz, including the investigation of a wide range of problems, among which are the relation between "general intelligence" and the ability to perceive consonance differences; the fusion of non-musical intervals; the determination of suitable criteria for consonance judgments; and the effects of repetition on comparative judgments of certain consonant or dissonant intervals. A few of the outstanding results of these studies will now be indicated, leaving the detailed consideration of them for later discussion. In the first place, low correlations have been found between "general intelligence" test scores and comparative judgments of the consonance of intervals. Second, the investigations of Guernsey show that certain criteria are more useful than others in making consonance judgments. Third, it has been shown that certain consonant intervals having simple vibration ratios can be slightly mistuned, thus making their ratios very complex, without affecting their consonance values. Fourth, certain data have been interpreted by some psychologists to mean that our general preference for consonant intervals is a matter of association. Fifth, almost all investigations of consonance perception have shown that some relationship exists between pleasantness and consonance, although the nature of this relationship has not been precisely determined.

Notwithstanding the considerable amount of experimentation in this field, little has been done in the way of contributing to an understanding of the fundamental nature of consonance perception. The contradictory results secured by various investigators renders problematical the value of a considerable amount of data which, otherwise, would serve as a certain basis upon which future research might be established. This unsatisfactory status of the work dealing with consonance perception is a direct result of the failure of investigators to obtain consistent judgments of relative consonance, and so long as this condition obtains no advance in our knowledge of this phenomenon is possible. This failure to secure consistent results indicates a corresponding failure to control the experimental conditions which exist when

the judgments are made. That is to say, the inconsistency of consonance judgments is doubtless due to the influence of a complexity of factors whose precise effects can be determined only by careful analytical experimentation. The mere repetition of consonance " tests "—as has too often been the case—cannot be expected to yield data of the sort essential to a genuine scientific analysis of the problem. The gross " score " of a subject on a consonance " test " probably is of little scientific value, not merely because such scores have been found to be unreliable, but more especially because the several responses of the subject, which are generalized into a single numerical value, are very probably diversely conditioned.

The present study aims primarily to investigate the effects upon consonance judgments of three sets of conditions: (1) the difficulty of the comparisons; (2) affective-tone; (3) the criterion or criteria used. An examination of the various experimental studies which have appeared during recent years shows that these three sets of conditions have constituted formidable difficulties for almost every investigator dealing with the problem of consonance perception. As early as 1918 both Malmberg and Gaw recognized that in attempting to measure consonance discriminability some allowance should be made, in the method of scoring, for the differences in difficulty which obtain between different pairs of intervals. However, aside from its importance for the practical problem of consonance testing, the matter of paired-interval difficulty has, for the most part, been ignored by writers in this field. This neglect has been unfortunate, since it would seem that we have here an important factor capable of influencing both the " accuracy " and the consistency of consonance discrimination. None of the comparisons in a series of paired-interval judgments is unconditioned by the setting in which it occurs. An " independent " judgment made purely in terms of the relative consonance of any two intervals in the series probably does not usually occur. Aside from factors such as progression which tend to stimulate affective judgments, the difficulty of a given paired-interval comparison can so disturb a subject as to interfere seriously with succeeding judgments. A major prob-

lem in the present study will be the experimental analysis of such effects.

As previously stated, almost all investigators of consonance perception have expressed the opinion that affective-tone influences the judgment of subjects. And since pleasantness is not generally held to be synonymous with consonance the operation of this factor has usually been regarded as unfortunate. Thus, for example, the consonance test studies of Malmberg and Gaw are admittedly open to the criticism that the judgments secured for them were "unduly" influenced by musical agreeableness. In fact, so common has become the notion that so-called consonance judgments are influenced by affective-tone that some thinkers have even gone so far as to question the possibility of non-emotional, cognitive response to such paired auditory stimuli. Still other writers, *e.g.*, Heinlein and Guernsey, while admitting the influence of affective-tone upon "consonance" judgments apparently regard the distinction between consonance and pleasantness as of little or no importance. That subjects should tend to be influenced by what is perhaps the most obvious characteristic of an auditory stimulus such as a musical interval, appears to be a logical supposition. However, notwithstanding the opinions of various writers on the subject to this effect, there exists no conclusive evidence as to the influence of affective-tone upon consonance discrimination. Heinlein's work, which consisted mainly in studying certain types of judgment reversals incident to changes in order of presentation of intervals within a pair, undoubtedly constitutes an important step toward the solution of the foregoing problem. However, his failure to secure both "consonance" and "preference" judgments, for the same paired-intervals, from his subjects renders his conclusions largely hypothetical. In view of the importance of this problem the present study has included certain experiments which, collectively, may be regarded as crucial for the point in question. That is to say, an analysis has been made of the relevant results obtained for four "consonance" and four "preference" tests given to 36 subjects. A general study has been made of the relation between consonance judgments and preference judgments by comparing the averages and reliabilities of the two

series of tests, and also by intercorrelating the scores made on the two types of tests. In addition to these general comparisons a detailed analysis of the effects of order of presentation of intervals upon consonance discrimination has been made. This has been accomplished by comparing the results on certain pairs of intervals when the subjects were instructed to disregard affective-tone and to make detached judgments on the basis of relative consonance with those secured for the same pairs when affective-tone was made the basis of the decisions.

Although various writers have called attention to the necessity of providing subjects with the proper criteria of consonance upon which to base their judgments, few have made this problem the subject of experimental study. Malmberg, one of the few investigators to give this problem serious consideration, made what was probably his most important contribution to experimental technique in the study of comparative judgments of consonance by defining three criteria—*blending, smoothness,* and *purity*—for his subjects, and instructing them to use only one of the three (the most appropriate) in any given comparison. However, this writer was concerned merely with securing consistency between the standard and the empirical orders of ranking intervals within the Octave c′ c″, and consequently made no attempt to determine the possible bearing of the various criteria (or the manner in which they are applied) upon the accuracy and consistency of consonance judgments. More recently, Guernsey has conducted an investigation which had for its aim an evaluation of fusion, smoothness, and affective-tone as criteria of consonance. As a result of this study she concluded that pleasantness and unpleasantness are the most legitimate criteria of consonance. However, several important questions arise in connection with the present problem which neither of the above studies answers satisfactorily. Is the attempt to use *blending, smoothness, fusion,* and *purity,* collectively, as consonance criteria conducive to "inaccuracy" and inconsistency in judging relative consonance? The criteria *blending,* and *smoothness* are regularly referred to in consonance test directions as though they were synonymous, yet the attempt to apply them to certain pairs of intervals shows that they frequently lead to divergent judg-

ments. The Major Seventh is obviously smoother than the Minor Second, yet the two tones which constitute the latter interval seem closer together and hence more nearly blend than do those constituting the Major Seventh. Thus, whether the Major Seventh is regarded as more consonant or less consonant than the Minor Second is partly dependent upon the criterion upon which the judgment is based. This is perhaps also true of other paired-intervals such as the Major Third and the Major Sixth. Furthermore, the writer has frequently noted that when the Major Third is compared with the Perfect Fourth on the basis of relative purity the latter interval is rather uniformly regarded as the more consonant of the two, whereas when blending is the criterion employed there is a tendency to judge the Major Third as the more consonant. Such considerations as the foregoing indicate that the various criteria cannot be regarded as synonymous, and suggest that much of the traditional "inaccuracy" and inconsistency of consonance judgments is due to the attempt on the part of subjects to judge relative consonance by using, collectively, criteria which severally give rise to divergent results. If this be true, then the further question arises as to whether there is any method of applying criteria that will materially lessen this inaccuracy and inconsistency? The answers to these questions—which the present study proposes to furnish—are of vital importance for any study of consonance perception, since consistent judgments of relative consonance are essential to progress towards an understanding of this complexly conditioned phenomenon.

In order to envisage the problem of consonance discrimination in its theoretical and experimental ramifications, the account of the present investigation will be preceded by an historical introduction, concerned with the evolution of the problem of consonance from Pythagoras to the present. In a sense, the development of conceptions and methods of studying this phenomenon symbolizes the general course of psychology from antiquity: mysticism, rationalism, and "scientific" elementarism. Although it is probably impossible to differentiate sharply between these various stages in the development of the present problem, they are perhaps represented approximately in the work of Pythagoras,

Euler, and Helmholtz respectively. And in the case of the present study it will be shown subsequently—without any attempt to formulate any specific theory of consonance—that consonance perception is not a simple all-or-none sensory response. Due to the influence of resolution, progression, and other factors to be discussed later, subjects probably do not react to a single paired-interval at a time in discriminating relative consonance, but rather to a 'pattern' or 'whole' of which the particular combination is only one part. In setting forth this historical development of conceptions dealing with consonance perception the attempt will be made to show that the various theories of consonance cannot be regarded as satisfactory because, in general, the kind of auditory phenomenon which is implied in the usual meaning of the term "consonance" has no actual existence. Criticisms of the various theories have frequently appeared in print but these have regularly been guilty of what seems to be, in the light of the present study, the most serious defect of the very "explanations" which they propose to invalidate, namely, over-simplification. That is to say, most of the criticisms of the various theories of consonance have made the same fundamental assumption as have the theories which they have undertaken to criticize. After having tacitly accepted the view of consonance as an elementary, all-or-none sensory phenomenon they have proceeded to criticize the traditional theories for their failure to account for this "fiction." In so doing they have neglected to call attention to a fundamental misconception whose earlier consideration would have made unnecessary the present treatment of the traditional theories of consonance. Furthermore, as will be shown presently, the projection of this over-simplified point of view into the experimental studies dealing with consonance perception has given rise to several contradictory results which do not contribute to an understanding of the phenomenon in question. In short, the attempt will be made to show that the history of both the theoretical and the experimental studies bearing on consonance perception reveals that insight into the fundamental nature of this complexly conditioned phenomenon can be gained only by means of a careful analytical study of the factors which condition it.

## CHAPTER II

### Historical Survey

It is probable that the early Greeks first subjected the phenomenon of consonance to rational consideration. They regarded the intervals of the octave, the fifth and the fourth as consonances, and referred to the latter as a "mixing of two things so that they are blended and form a compound." Pole (21, p. 109) calls attention to the fact that although part music was unknown to this age, it appears from the writings of Euclid that the Greeks possessed some notion of the harmonic relations of the principal consonant intervals. Euclid alludes to the consonant blending of a higher with a lower tone in the cases of the octave, the fifth and the fourth, as distinguished from all other intervals. Notwithstanding the fact that their criteria of consonance are largely unknown, it seems probable that for them consonant intervals were uniformly those which were pleasing. Such an hypothesis seems warranted in consideration of the nature of Greek aesthetics, since according to the latter simplicity was an important element of beauty. It was to be expected, then, that they should have regarded the simpler consonances as the more pleasing intervals. Thus it happened that their notion of what was harmonically pleasing coincided with what we today regard as consonant. In more recent years, however, many of the more complex forms of auditory experience have also come to be considered aesthetically pleasing. Thus certain relatively complex and dissonant intervals have come to be regarded as of aesthetic value within certain settings. Because of this change in practice, the pleasing is no longer identical with the consonant, as it was in the case of the early Greeks, and accordingly, affective-tone has ceased to be a safe criterion of consonance. However, it will be seen from the following discussion that this distinction has not been uniformly maintained by the various writers on the subject.

## 1. *Theories of Consonance*

As early as 600 B.C. it was recognized that those intervals which have the more simple vibration-ratios are the more consonant, and accordingly the first attempts to explain consonance were based chiefly upon the assumed importance of this fact. Later, however, it was discovered that it is possible to mistune slightly certain consonant intervals, thus making their ratios very complex, without affecting their degrees of consonance. For some thinkers this discovery constituted a direct contradiction of the foregoing formulation, which had come to be accepted as an established fact, and hence seemed to demand a radical change in approach to the problem. On the other hand, Lipps denied that these apparent exceptions were of cardinal importance, holding that they could be adequately accounted for by recourse to Weber's law. There are other facts which are closely related to the phenomenon of consonance which some thinkers regard as of paramount importance for its explanation. In the first place, it is known that certain tones when sounded simultaneously give rise to beats and that these impart a rough jarring effect to the tone. The fact that beats are generally present when tones are dissonant and absent when tones are consonant has led to the belief that the two phenomena are necessarily related. Secondly, the pleasingness of an interval and its consonance coincide sufficiently to give rise to the notion that, although not identical, they are related in some manner. Thirdly, the history of music shows that intervals which were once regarded as inharmonious have come in more recent years to be looked upon as suited to musical treatment. Some psychologists hold this fact to be of great significance for the explanation of consonance, since it apparently indicates that the perception of consonance is the result of an evolutionary process. Regardless of their importance, however, differential emphasis upon these several sets of facts has given rise to various types of theories which regard consonance from at least four points of view: (1) the mathematical factors involved; (2) the physical and physiological factors; (3) the psychological factors; and (4) the genetic factors. The

following summary[1] of theories will serve to show how various investigators have attempted to account for such of the above facts as they have deemed pertinent.

As Guernsey (6, p. 175) points out, Euler and Schopenhauer are ordinarily classified as exponents of the purely Pythagorean mathematical basis of harmony. Leibniz (15) was the first to point out that the mind does not really analyze or perceive the actual number or the numerical regularity of the vibration frequencies in the intervals. In order, however, to account for the fact that the consonance of an interval seems in some way dependent upon the simplicity of its vibration-ratio, Leibniz resorted to his hypothesis of the "unconscious mind" and came to the conclusion that the latter calculated the ratios of the vibration frequencies. Euler (10), who was in essential agreement with Leibniz, attempted in a more serious and more scientific manner to show the relations of consonance to whole numbers. He based his theory upon the assumption that we are pleased with everything in which we can detect a certain amount of perfection; consequently, a combination of tones will please us when we can discover the law of their arrangement. Since agreeableness is directly proportional to the ease with which the order can be discerned, it follows that the combination of two tones will please us the more the smaller the two numbers by which the ratios of their periods of vibration can be expressed. Schopenhauer (24), while recognizing the factual basis of Euler's view, held that consonances and dissonances portray the movements of the human will in its essential feelings of satisfaction and dissatisfaction. Such theories are probably of little value, since they are tautological rather than explanatory, and are so conceived as to be impossible of scientific attack.

The theory of Lipps has been variously classified by different thinkers, but since it is admittedly a return to the older point of view as set forth by Euler it seems logical to place it in the cate-

[1] Discussions of the various theories have appeared so often in print that the present review will confine itself to a brief statement and criticism of the more important ones. In presenting this summary the present writer acknowledges indebtedness to Malmberg's historical investigation (17) and to Helmholtz's (10, p. 229) description of the theory of Euler.

gory of mathematical explanations. Lipps (**16**, p. 142) holds that the essential element in consonance is that of agreement, not *any* sort of agreement, but a feeling of pleasurable, inner consistency or unanimity. According to Lipps this feeling of unity is due to the presence of a common rhythm between the two tones constituting an interval. For example, we designate the " rhythm " of the sequence of 100 vibrations per second as " rhythm of 100," and similarly, we designate the rhythm of the sequence of 200 vibrations per second as " rhythm of 200." Now, these two tones have a common rhythm of 100 which serves to bind them together. By the same token that the psychic excitation discharged by the sequences of physical vibrations is conditioned by the nature of these sequences, so these vibration-rhythms, which constitute the essence of the various vibration-sequences, are held to " resound " in some way or other in the corresponding psychic excitations. The important thing here, however, is not that the vibration-rhythms themselves are preserved in the processes of sensation but rather the preservation of these relations in the psychic processes. Further, because of the fact that the fundamental rhythm includes in itself the two tones in an ever higher degree, dependent on the simplicity of the vibration-ratio and hence of the ratio of the rhythms in the separate immediate experiences of tone, the simpler the vibration-ratios, the higher the degree of unification and the greater the degree of consonance. This theory, which is a combination of the views of Euler and Leibniz, is open to the same criticisms which were raised with respect to the latter theories: it is tautological rather than explanatory; it seeks to explain the conscious by means of the unconscious; and it fails to account for the fact that it is possible to mistune slightly certain consonant intervals having simple ratios without affecting their consonance values. Lipps has attempted to meet the first two objections by stating that vibration-ratio is decisive for consonance in a manner precisely analogous to that in which vibration-rate is decisive for pitch. The third objection is held to be duly answered by recourse to Weber's law. However, the validity of these defenses is problematical. In the first place, the notion that the

"sense" of consonance is perfectly analogous to the sense of pitch is contradicted by the results of almost all recent studies which have indicated that consonance perception is a comparatively complex process. Secondly, the variability of consonance measures suggests rather strongly that Weber's law does not hold for the "sense" of consonance. However, as far as the present study is concerned, the most serious defect in Lipps' theory is that it does not lend itself readily to experimental attack, since the "micropsychic" rhythms are non-verifiable entities[1] deduced from the conscious experience which they are intended to explain.

The theories of Helmholtz and Krueger may be regarded as examples of the second type of explanation, *i.e.,* as emphasizing the physical and physiological factors involved in consonance. The theory of Helmholtz constitutes the first attempt to account for the phenomenon of consonance on a purely physiological basis. According to him (**10**), dissonance is due to beats taking place between either the tones themselves or their upper harmonics or the differential tones to which they give rise. A clearly marked consonance occurs when the ratio of the two tones is such that there are no beats, but when a slight change in the ratio gives rise to beating. By reason of the intermittent and discontinuous stimulation which these beats afford, "the nerves of hearing feel these" (**10**, p. 226) as rough and unpleasant. This theory constitutes an advance over the older point of view in that its criterion of consonance is a fairly obvious characteristic of the sensations themselves, and also since the mechanisms alleged to be involved in the perception of consonance are definitely perceivable structures rather than abstract mentalistic entities. However, several considerations serve to show the inadequacy of Helmholtz's theory: first, it offers an explanation of dissonance rather than of consonance;[2] second, it fails to account for instances in which

[1] However, in fairness to Lipps it is perhaps well to state that he thought of "micro-psychic" rhythms in about the same manner that we regard genes, atoms, electrons, and all other analogous constructions—*i.e.,* as mere hypothetical constructions to account for certain facts of experience.

[2] Peterson (**20**, p. 20) holds that this criticism is not entirely valid, since Helmholtz mentioned coincident partials as well as absence of beats as a condition of consonance. However, it should be noted that Lipps (**16**, p. 167) considered this claim made by Helmholtz and rightly rejected it as begging

dissonance is heard without perceptible roughness; and third, in emphasizing but a single criterion it is at variance with the results of recent studies (**14**, **17**) which have shown consonance to be of too great a complexity to be judged on the basis of any single criterion.

Krueger's theory (**12**) really constitutes a supplement to that of Helmholtz. Whereas the latter emphasized chiefly the effects of overtones, Krueger attributes dissonance to the presence of beats generated by combination tones, more especially by difference tones. According to Krueger the five difference tones of two simultaneous tones are to be calculated by subtracting the lowest from the next lowest tone (*e.g.,* 4 : 5 > 1, 3, 2, 1, 0). Krueger emphasizes particularly the characteristic difference tone whose ratio always corresponds to the number 1. In the case of pure consonances this tone is due to the identification of two or more difference tones. For example, the interval 200 : 300 has at least three difference tones of 100 vd. each. However, since these three difference tones are identified with each other or with the "characteristic" difference tone the interval is consonant. If by slightly mistuning the interval we obtain the ratio 200 : 307, the difference tones become 107, 93, 14, 79, and 65. These tones lie close together and hence give rise to beats which occasion dissonance. On the other hand, consonances are always free from difference tone beats, and contain only perfect unisons, since the difference tones either coincide or else are a third apart. Although this theory proposes a scientific rather than a mere verbal solution of the problem it is open to several criticisms:[1] first, it offers an explanation of dissonance rather than consonance; second, it fails to account for the dissonance of such an interval as 8 : 11 whose difference tones are 3, 5, 2, 1, 1; third, it fails to account for the fact that certain auditory disturbances may be artificially produced coincident with a consonant chord

the question since the klang-affinities as consonances were used by Helmholtz to explain consonance, whereas they themselves need to be explained.

[1] For a more detailed treatment of the theories of Helmholtz and Krueger, as well as those of Stumpf and Wundt, the reader should consult Lipps' study (**16**), which contains perhaps the first thoroughgoing criticism of these views.

such as c–e–g without altering the consonance value of the letter; and fourth, in emphasizing but a single factor it seems, in the light of recent studies (14, 17), to be guilty of oversimplification. Hence Krueger's theory, as well as that of Helmholtz, must be regarded as a view which emphasizes a single important factor influencing consonance perception, rather than as an adequate explanation of consonance.

The theories of Stumpf and Wundt are similar in that both are essentially "psychological" theories. For Stumpf (29) "Verschmelzung" or fusion is the distinguishing criterion of the degree of consonance which must be sought in conscious experience and which is due to the perception of a quality inherent in the tones themselves. According to Stumpf, fusion is not mere unanalyzability; rather it is the qualitative unity that persists in a chord after its indistinguishability has been superseded by a consciousness of separate intervals. This residual qualitative unity is held to be an original relation, like a sensation of color, and is not to be referred to anything more ultimate psychologically. Notwithstanding the importance of Stumpf's work on the problem of tonal fusion several objections to his theory remain unanswered: first, as pointed out by Moore (18), the distinction between the unity which disappears with analysis and the unity which remains after analysis is a dubious one which Stumpf himself has maintained with but varying degrees of success; second, Stumpf's claim that fusion is independent of absolute pitch, of the relative intensities, and of the *timbre* of the primaries of the interval is contradicted by the experiments of recent investigators such as Peterson (20); third, in drawing a sharp line of demarcation between consonance and dissonance it assumes a simplicity in consonance and dissonance which is at variance with the results of the experimental studies already referred to (14, 17); and fourth, a recent investigation by Guernsey (6) has shown that "fusion," instead of being the sole criterion of consonance, is merely a highly ambiguous term which, when made the basis of consonance judgments, gives rise to inaccuracies and inconsistencies on the part of the subjects.

Wundt (31) was one of the few thinkers to realize that con-

sonance is not a simple, uniformly conditioned phenomenon. He held that since we do not always feel the same in the presence of chords and sequences of tones which we designate as consonant there must be a plurality of factors involved in consonance. Accordingly, he recognized three criteria of consonance: (1) the relatively narrow unity of fusion, (2) distinctness of tonal fusion, and (3) dominating tonal element. In the final analysis, however, consonance is an act of the apperceptive faculty of the mind which synthesizes the tones into a unity. In holding that consonance is a complex phenomenon conditioned by several factors Wundt has struck a modern note which has received increasing verification by recent investigators. However, his theory is of little scientific value, since in holding that consonance is the result of a synthesis effected by the 'apperceptive faculty of the mind' he offers a purely verbal explanation which is impossible either of verification or of disproof.

In contrast to the older conceptions of consonance, Malmberg (17) holds that the perception of consonance is a cognitive process, involving the factors *blending, smoothness* and *purity*.[1] Furthermore, these criteria of consonance apply with varying degrees of appropriateness to different pairs of intervals. Certain combinations can best be judged on the basis of *blending,* for others the determining criterion is *smoothness,* while for still others it is *purity*. Although this view is merely a hypothetical statement of the conditions of consonance perception rather than its explanation, it seems to constitute a step in the right direction, in that it tacitly recognizes the complexity of consonance and takes as its criteria fairly obvious characteristics of the tones themselves.

The genetic theories of Ogden and Moore represent more recent attempts to account for the phenomenon of consonance. According to the former (19, p. 142), consonances and dissonances are the result of congenital dispositions developed during the history of the race. The frequent hearing of certain combinations of tones is held to modify the auditory mechanism in such a way that these combinations become more consonant, and

[1] In one place Malmberg refers to "fusion," but this criterion is eliminated in his later investigations.

this modification is presumed to be inherited by succeeding generations. This theory merits little consideration since it is contrary to all known facts of inheritance, and since, in the final analysis, it explains nothing.

The theory of Moore (**18**, p. 62) differs very little from that of Ogden, save in its omission of the inheritance factor. According to Moore, consonance is a special case of the adjustment of the nerves to outer relations. "The nervous system, by a form of activity that tends with each repetition to become more simple and economical, gradually affects the synthesis of more complex physical relations." The evidence which Moore cites in defense of this view is taken principally from the history of music. Examination of the latter shows that since the beginning of the eleventh century music has become increasingly complex, due chiefly to the gradual introduction of dissonances. Thus, certain intervals which were originally regarded as inharmonious and therefore not suited to musical treatment have finally come to be accepted as indispensable elements in musical composition. This fact is interpreted by Moore to mean that consonance has accordingly undergone a gradual evolution, and that in several instances intervals which were once looked upon as dissonant have gradually come to be regarded as consonant. This view marks a departure from the more static conceptions of consonance, and has accordingly gained widespread attention among psychologists. Because of the importance which has been attached to this theory, a critical examination of it would seem appropriate.

Moore's theory seems to rest chiefly upon a misinterpretation of the significance of the history of music for the problem of consonance, and upon the tacit assumption that the pleasing and the aesthetic are identical. It will be recalled by those conversant with the history of music, that the introduction of dissonances into music came about mainly as the by-product of contrapuntal music. Composers found that in combining different melodies discords were unavoidable at certain points. At first these incidental clashes occurred for the reason that they could not be avoided without handicapping the composer, and hence were tolerated rather than sanctioned. At a later period, however,

characterized by a higher degree of aesthetic sophistication, they were introduced for aesthetic effects, although their use was restricted by certain rules. The use of dissonances probably reached its artistic climax with Wagner and Richard Strauss. For these men the art of music meant far more than it did to the early Greeks; it was a means for the portrayal of the entire gamut of human emotions—for hate as well as for love, for sorrow as well as joy, and for conflict as well as peace. Manifestly, no arrangement of simple consonances could possibly be adequate to such a portrayal, and the more modern composers have accordingly made extensive use of dissonances as a means to aesthetic enrichment. It is obvious that this practice does not mean that intervals or chords which were formerly regarded as dissonant have become consonant. As just stated, whatever dissonances or discords originally occurred in music were the unavoidable by-products which resulted from combining two or more melodies. Modern composers, however, have made rather free use of discords for their aesthetic effects within certain settings. The value of these discords for the securing of these effects is absolutely dependent upon their being heard not as consonances but as dissonances. Psychology, in studying the fundamental conditions of consonance perception is interested neither in the pleasingness nor in the aesthetic value of an interval within such a setting, but rather in the relative consonance of an interval in comparative isolation. The problem of the relation of consonances and dissonances to the aesthetic effects of music undoubtedly has its place in psychology, but it should not be confused with the problem of consonance *per se*. On the other hand, an interval in isolation has no meaning for music. Thus, because of this fundamental difference in interests, it is a mistake to attempt to make the history of the *art* of music the basis for a *scientific* explanation of consonance.

In addition to this error of interpretation, Moore's position seems susceptible of more specific criticism. In the first place, Moore speaks of the adjustment of the nerves to outer conditions. It is doubtful, however, whether repetition would modify any part of the mechanism involved in hearing to the degree to which

Moore's theory would seem to require. What can Moore possibly mean when he assumes that an interval such as the Minor Second will eventually come to be regarded as a unity? In order for such a synthesis to be effected some drastic kind of central or peripheral modification of the auditory mechanism would be necessary, and, as just pointed out, such a change is very improbable. In advocating this theory Moore seems to lose sight of the fact that consonance discrimination rests primarily, not on habit or custom, but on the ability of the individual to perceive that an interval is composed of two tones of different pitch. That is, before an individual can perceive the *relation* between two tones he must first be able to perceive the two tones, although he may not always be explicitly aware of their independent existence. This implies the ability to discern differences in pitch, otherwise the two tones would be heard as one. Hence, in order for such an interval as the Minor Second to come finally to be regarded as consonant one of two things would seem to be necessary: either there must be a progressive decrease in the ability of the individual to perceive differences in pitch, or else these pitch differences must in some manner fail to function as cues for the appreciation of tonal relationships. Secondly, the theory seems to ignore entirely the nature of the physical factors which Helmholtz and others have held to condition the perception of consonance. Thirdly, in holding that the frequent hearing of an interval tends somehow to unify it, Moore overlooks the fact that it is the trained ear which distinguishes between a note and its octave when they are sounded together, whereas the untrained individual often confuses the two and is unable to distinguish two tones even when they are called to his attention. This is just the opposite of what the above theory would lead one to expect.

It would seem, then, that Moore's theory deserves to be placed in the same category as that of Ogden—as one which makes impossible demands with respect to modification of the mechanisms involved in hearing. The fact that the consonant and the harmonious were originally identical has been mainly responsible for the mistaken notion that the history of harmony is also the history of consonance. Apparently, this confusion could

have been avoided by noting the fact that harmony and consonance began to diverge with the introduction of contrapuntal music, and that the increased discrepancy which obtains today is attributable, chiefly, to the development of a broader conception with respect to what may be regarded as aesthetically valuable.

The foregoing criticisms of the various theories indicate that, for the most part, the latter are merely ideal constructions with little foundation in fact. They have tended to regard consonance as a comparatively simple and uniformly conditioned phenomenon to be clearly distinguished from dissonance, and have been couched in an abstract terminology which has rendered them unsuited to scientific attack. Even the more truly scientific explanations such as those of Helmholtz and Krueger were seen to be oversimplifications, since each rests upon one major fact, the inadequacy of which is made manifest by the failure to account for the perception of dissonance in the absence of beats. Although avoiding the abstractness and oversimplification of the earlier views, the recent attempts to regard consonance as a genetic process have failed to provide a satisfactory explanation. They have circumvented rather than met the real problem and in doing this have become involved in difficulties as serious as those from which they have apparently escaped. In short, the attempts to formulate theories of consonance independent of empirical study of its factual basis have resulted in unprofitable verbalisms, vagaries, oversimplifications and inconsistencies. In the words of Bertrand Russell, it is becoming increasingly evident that what this problem demands is not further *a priori* speculation but "the substitution of piecemeal, detailed, and verifiable results for large untested generalities." This needed shift in emphasis is being gradually effected. The many experimental studies undertaken in this field during the last two decades are indicative of the change in the method of attack upon this problem. The results of some of the more important of these will now be presented.

## 2. *Experimental Studies*

One of the earliest experimental studies was that of Emerson (2) relative to the problem of interval preferences.

Working with reduced intervals, he claimed to have demonstrated that the apparently natural demand for the tone combinations which give fusion (or consonance) can be inhibited during the listening to non-musical combinations, as soon as a short training in miniature intervals changes the acoustical perspective. Thus, Emerson makes the general preference for consonances a matter of association, holding that the fact that our tone-consciousness has been trained in our musical tone-relations is responsible for our apparently natural preferences.

Several years later Moore (**18**), continuing this same type of investigation, reported the historical and experimental study already mentioned, which seems to support the view that dissonance is the result of strangeness, and that any interval may finally come to be regarded as consonant if heard a sufficient number of times. Moore's historical data based upon the history of music have already been discussed. His experimental evidence is based upon experiments performed on nine subjects. The problem was to find out whether or not the prolongation and repetition of certain intervals produced anything which might be interpreted as a change in their degrees of consonance, such a change being indicated by the amount of change in the acceptability of an interval as a parallel. The general method was that of paired comparison, with each judgment graded by the subject according to the degree of his preference. At the beginning of each hour of experiment a table of graded comparisons of the four parallel intervals in question was constructed, to which was added the minor ninth in order to give a wider range of comparison. This table, which served as the standard for the day, was made up as follows: each interval, played consecutively in a passage of seven parallels, with c′, d′, e′, f′, e′, d′, c′, as the fundamentals, was compared with each other one played similarly. In each case the subject was asked not only to state his preference between the two passages, but also to grade the strength of his preference according to a scale A, B, C, D, E. After thus obtaining the standard for the day the particular interval for that day's investigation was studied, either by the method of prolongation or of repetition. In the case of the former method

a passage of parallels, involving the interval to be studied, was played, each interval being sustained one minute. This procedure was continued for five minutes. Immediately after this, new preference judgments were called for, in which the comparison was made only between the particular interval under consideration and the other parallels. The method of repetition differed from that of prolongation in that each period was occupied not with sustaining intervals, but with playing repeatedly an entire melody in parallel thirds, fifths, minor or major sevenths, as the case might be. Both of these methods showed certain characteristic tendencies: the third lost rapidly; the minor seventh gained equally rapidly; the fifth maintained a fairly constant level; the major seventh rose in value, but less rapidly than the minor seventh. That is, repetition of a dissonance tends to raise the consonance value or to lessen the dissonance value of that combination. Thus, Moore concludes that consonance is the result of the adjustment of the nerves of hearing to frequent repetition of an interval. However, the value of Moore's results are dependent upon the correctness of his assumption that the consonance of an interval is inversely proportional to its acceptability as a parallel interval. This really amounts to making affective-tone the basis for consonance judgments, whereas it is generally assumed that the perception of consonance is primarily a cognitive process.

One of the principal problems which has arisen in relation to the general study of consonance is that of the measurement of individual differences in ability to judge the relative consonance of various intervals. This type of study had its beginning at the State University of Iowa in a series of investigations by Seashore and several of his students. As early as 1910 Seashore published a preliminary report (25) concerning the measurement of pitch discrimination, and in 1915 an article by the same author appeared which dealt with the psychology of individual musical talent, classifying the principal measurable traits. The latter article showed how data resulting from the tests could be reduced to a "common denominator" or norm and suggested the meaning and use of such a procedure.

One of the first attempts to "measure" consonance discrimination was made by Malmberg (17) at the University of Iowa. This problem necessitated two investigations, the first of which had for its aim the standardization of a test for the measurement of the "sense" of consonance. The method used in this preliminary investigation was as follows: first, the determination, by a survey of the historical theories of consonance, of the factors which enter into its perception; second, with these factors as a basis, the intervals were ranked according to their relative degrees of consonance or dissonance; and third, the evaluation of an individual's ability to compare the consonance values of intervals in terms of this ranking. As a result of this investigation, Malmberg found the constant criteria of consonance to be *blending, smoothness, fusion* and *purity*. Upon the completion of this preliminary investigation, a second investigation was undertaken with three ends in view: first, to secure measurements of individual differences; second, to establish norms; and third, to test these measurements under controlled conditions. As a result of this second investigation, Malmberg found a satisfactory agreement between the standard order and the empirical order of degrees of consonance of the intervals within the Octave c′, c″. The deviations occurred chiefly in the consonances; and these were due mainly to the fact that the Major Third was unduly preferred in the empirical rankings. From this study Malmberg concluded: first, that the historical failure to reach an agreement as to the relative consonance values of the various intervals was due principally to a disagreement as to what constituted consonance; second, that the perception of consonance was a cognitive process dependent upon an elemental sense;[1] and third, that the constant factors involved in the perception of consonance were *blending, smoothness* and *purity*.

Gaw (4, p. 141) continued the type of investigation started by Malmberg, effecting a revision of the latter's sixty-six unit test by the elimination of certain pairs of intervals which seemed to

[1] Malmberg says (17, p. 131), "The perception of consonance is elemental in a secondary sense in so far as it is based rather on the elemental capacities for pitch discrimination and tonal memory than on acquired musical ability or training."

be either too easy or too difficult. The result of this revision, however, was an eleven unit test which she regarded as too easy for use in testing the abilities of individuals to discriminate consonance differences.

The notion of measuring musical talent which had received impetus through the investigations of Malmberg and Gaw reached its practical climax in the publication of Seashore's Measures of Musical Talent by the Columbia Phonograph Co. These records include six tests, each of which purports to measure a native elementary " sense " or " capacity " which is held to be essential to musical appreciation and performance. This series of tests includes: Test 1, Sense of Pitch; Test 2, Sense of Intensity; Test 3, Sense of Time; Test 4, Sense of Consonance; Test 5, Tonal Memory; and Test 6, Sense of Rhythm. The Consonance Test is composed of fifty units, as shown in table 11 *(infra)*. In devising this test Seashore relied to a great extent upon Malmberg's study, making *blending, smoothness* and *fusion* the criteria of consonance. However, Seashore deviated slightly from the procedure of Malmberg in that he substituted the criterion *fusion* for *purity*. In a recent article (27, p. 181) Seashore states that this substitution was made because more of the judgments called for in his test seemed to be determined by *fusion* than by *purity*.

Seashore's tests for measuring native musical capacities have been rather severely criticized. Probably the most unsatisfactory one of these is the consonance test, which is based upon the assumption that consonance perception is due to a measurable elementary sensory capacity which differs innately in different individuals. This notion has not met with general acceptance. In the first place, many psychologists have failed to subscribe to the belief in the possibility of devising a valid and reliable test of this kind. The notion that it is possible, within a few minutes, to measure a complex capacity which seems to vary with so many conditions, and which manifestly is not a " sense " in the true meaning of the term has met with much opposition. In this connection it must be remembered that a " sense " implies a

sense-organ, and that there is no known organ for the perception of consonance.

One of the principal critics of the Seashore Consonance Test is Heinlein, who, by reason of the increasing use made of dissonances by composers of modern music, has even gone so far as to intimate (8, p. 433) that although the test were both valid and reliable (as a measure of consonance perception) it would still be useless. In 1925 Heinlein undertook an experimental investigation (8) for the purpose of ascertaining the extent of the influence of harmonic principles and of the various laws of musical progression upon judgment in the paired-interval comparison method employed by Seashore in the consonance test. The Seashore test was given to a group of 35 subjects and then the test was repeated at the same sitting. Approximately two months later the same test was given twice to a group of thirty subjects (" fifteen had appreciable musical training and experience "). As a result of these two studies Heinlein concludes that the paired-interval comparison is an inadequate method for testing consonance. His results indicated that the presence of such unavoidable factors as progression, resolution, etc., tends to render the test invalid. Furthermore, musical training seemed to be productive of negative results since those with training made lower scores than the untrained and were less consistent in their judgments. Finally, Heinlein came to a position which is similar to that of Moore, namely, that the history of harmony would seem to warrant the assumption that such a test would have little value for modern music, and that it would have still less value for the music of the future.

Larson (14) has continued the line of investigation initiated by Heinlein and claims to have found that the Seashore Consonance Test is a reliable means for measuring the perception of consonance. She reports an investigation which was undertaken for the purpose of determining the reliability of the Seashore Consonance Test " when conditions are controlled as rigorously as possible and when the instructions are literally followed." The test was given as a group test and as an individual test, both trained and untrained subjects being used in the experiments.

In each instance the test was repeated. Five groups of subjects were used. Group A was composed of 132 students in an elementary psychology class, selected on the basis of scholarship. Group B consisted of 150 students in an unselected class in elementary psychology. Group C consisted of 35 musically trained subjects, and Group D of 35 untrained subjects. Group E consisted of 35 musically trained subjects who were tested individually, and Group F consisted of 35 untrained subjects tested individually. Comparing the results obtained in the first performance of the consonance test with those of the repetition several months later, Larson found a correlation coefficient of $.63 \pm .022$ for Group A, and of $.65 \pm .024$ for Group B. The mean of the trained subjects was "considerably higher" than that of the untrained subjects, both in individual and group tests. The trained subjects (Group E) showed greater constancy of error and greater general uniformity of judgment than the untrained subjects constituting Group F. On the basis of these results Larson concludes: first, that the paired-interval comparison method is adequate for testing consonance, and that the principle of harmonic progression is a negligible factor; second, that the test is reliable, both as a group test and as an individual test, when the instructions are followed; third, that affective judgments may be eliminated to a great extent under properly controlled conditions; and fourth, that musically trained subjects tend to make higher scores and are more consistent than those without such training.

Pratt (22) reports an investigation concerning quarter-tone music which is of special interest in connection with Moore's view that the future will see the invention of more dissonances for use in music. According to Pratt, intervals in order to be uniformly recognized as different in quality must be separated from each other by not less than a quarter-tone. Relative to the significance of this fact for music Pratt says,

"It would, therefore, be theoretically feasible to double our present chromatic scale so that it would comprise twenty-four quarter-tones, instead of twelve semi-tones. But any further subdivisions into eighth-tones and sixteenth-tones would be juggling with the physics of tuning instruments rather than catering to the psychological capacities of auditory discrimination. The

listener to music in which eighths and sixteenths were employed would be generally quite unaware of their presence."

This investigation seems to have rendered a signal service in calling attention to the differential limen as an element involved in the perception of consonance, a fact which investigations like that of Moore seemed to have ignored entirely.

Guthrie and Morrill (7) conducted an investigation concerning the fusion of non-musical intervals. Stern tone variators were used to produce 44 intervals ranging from perfect unison to an interval beyond the fifth. The subjects first ranked each interval as pleasant or unpleasant and later ranked them on the basis of consonance. The fact that the curves for pleasantness and consonance show high agreement raises the question whether the attempt to eliminate the affective factor from consonance judgments should be made.

Guernsey (6) reports an investigation in which she attempted to make an evaluation of fusion, smoothness and affective-tone as criteria of consonance. Using three types of subjects (musically untrained subjects, subjects with a moderate amount of musical training, and professional musicians) a series of tests was given in which fusion, smoothness and affective values were used as separate criteria. From this investigation Guernsey found that fusion and smoothness are inadequate criteria—tonal fusion being a sensorial rather than a perceptual phenomenon and smoothness being subject to too great a diversity in connotation in the mind of the listener. Further, she concludes from her results that pleasantness and unpleasantness are the most legitimate criteria of consonance. Relating these findings to certain tendencies in modern music, Guernsey finally concludes that " consonance is an aesthetic description, totally dynamic in nature, and not a scientifically determinable constant."

A study of much practical significance has recently been reported by Stanton (28). This investigation was concerned with the application of the Seashore Measures of Musical Talent in the Eastman School of Music, and was based upon the records of students entering the school from September, 1921, to the fall of 1924. After a period of three years there was found to be

such a close correspondence between the teachers' estimates and the test profiles that the faculty of the school decided by unanimous vote to exclude all candidates whose test grades were as low as D and E.

The foregoing survey of recent experimental studies shows that a variety of problems has engaged the attention of investigators. These have ranged from purely scientific studies of the conditions influencing the perception of consonance to more practical ones concerned with the devising of consonance "tests." The results of these various studies seem to indicate: first, that consonance is a complex phenomenon, and is conditioned by many factors, and that because of this fact no single criterion is an adequate basis for consonance discrimination; second, that notwithstanding the distinction between consonance and pleasantness they are related in some manner yet to be determined; third, that any theory, such as Moore's, which assumes a progressive conversion of dissonances into consonances must make allowance for the fact that the differential limen is a factor involved in the perception of consonance; and fourth, that probably the most important problem in this field is that of the low reliability of consonance judgments. The fact of "low reliability" would seem to indicate that consonance discrimination is influenced by complex conditions. Although this constitutes a difficulty from the point of view of experimental control, it is of theoretical significance in that it indicates that no explanation of consonance as a simple process can be regarded as adequate.

## CHAPTER III

### The Problems and the Experimental Methods of the Present Study

Investigators of " consonance " too often have assumed that it is a simple, rather uniformly conditioned perceptual phenomenon. This was seen to be true especially of the " theories " of consonance, which were shown in general to be rather futile exercises in the manipulation of concepts. The same criticism might well be made of those experimental studies which have presupposed a " sense " of consonance, with the implication that the reaction described as consonance discrimination (or comparison) involves merely the direct response of the sensory mechanisms to stimuli which could be compared in terms of a definite " linear " differential. In practical terms, this view leads naturally to a " test " of consonance discrimination, with a view to differentiating individuals on the basis of the type of sensibility involved. Unfortunately, the experimental work projected upon the basis of these over-simplified assumptions concerning the nature of " consonance " has not justified expectations. The " tests " have yielded quite inconsistent, " unreliable " results. Experimenters dealing with the concrete comparisons of intervals by individual subjects have been confronted with a rather bewildering array of factors which complicate the discriminations required. The abstract notion of a somewhat unitary process of " consonance perception " conditioned by constant relations among tonal stimuli has been found to be largely a delusion.

Such consequences were perhaps to have been expected. Psychologists, and other scientists as well, have found regularly that *a priori* verbal analyses in which processes were sharply defined, and logically classified—tend to evaporate when subjected to experimental observation. The clear-cut lines of demarcation disappear; the causal relations, which had been neatly generalized into precise laws, are found to be quite complex

and difficult to discover. It becomes necessary to resort to laborious experimentation, with a view to controlling and describing the effects of the factors which seem from observation to be related to the phenomenon in question. This seems to be precisely the present status of the problem of consonance. The inconsistency of comparative judgments of intervals, contradictory results and claims on the part of investigators due often to differences in criteria (*i.e.*, in experimental conditions), the persistence with which out-moded theories continue to saturate the literature, all these considerations suggest the necessity of unbiased experimental analysis. Accordingly, the major aim of the present study is to determine the effects of three important sets of conditions on the consonance judgments of rather homogeneous groups of college students. The factors to be studied are (1) the difficulty of the comparisons, (2) affective-tone, and (3) the criteria of "consonance" used as the basis of comparison.

The general experimental technique used was the paired-interval comparisons method employed by Seashore and others. Several variations were introduced in the arrangement of pairs of intervals, and the directions to the subjects. Throughout the description of methods and results the term "test" will be used as a convenient designation of a series of intervals presented as stimuli to the subjects. Neither the series of intervals on Seashore's record nor any other set devised for special experimental purposes should be regarded as "tests" in the real sense of this term—*i.e.*, with respect to intent and use. The general outline of the work will now be described, to be followed by a detailed account of each experiment. Seashore's Consonance Test, as embodied in the Columbia Phonograph Record No. 53001-D, was first given to a representative group of college students with directions designed to secure both "consonance" and "preference" judgments, for the purposes of providing data to be compared with results of previous investigators, and to be analyzed with respect to paired-interval difficulty and affective-tone. In order to check the consistency of the judgments and the effects of practice under these conditions the test was given

four times to the same group under each of the two sets of directions, *i.e.,* as a " consonance " and as a " preference " test. Secondly, in order to determine the influence of the difficulty of an experimental series upon consonance judgments, four other " tests " of varying difficulty were devised and given to another group of college students. Thirdly, in order to secure additional data with respect to the influence of different criteria upon consonance judgments two other tests were given to two additional groups of students. The first of these two tests called for the preferential use of the criteria *blending, smoothness* and *purity,* and was given for the purpose of determining whether the use of the above criteria in the manner indicated is conducive to reliability. The second test which was given to a different group, involved the use of each of the criteria *blending* and *smoothness* in separate series, and was given for the purpose of determining the reliability of subjects' judgments when the pairs of intervals were so arranged that a single criterion could be used at a time. As a check upon the comparative reliability of this test a " preliminary " test, similar to the one used in connection with our study of the " preferential use of the criteria *blending, smoothness* and *purity,"* was given to the group. In this instance, however, the two criteria *blending* and *smoothness* were used preferentially. A detailed description of the several sets of experiments, and of the conditions under which comparative judgments were made will now be given.

### 1. *The Seashore Consonance Test, with " consonance " directions*

The Seashore Consonance Test was given to 36 students from an introductory psychology class on November 4, 1929. This test, which is one of a battery of six tests devised for the purpose of measuring " native musical capacity," is composed of fifty pairs of intervals (see table 11 below) which have been recorded on a double-faced record by the Columbia Phonograph Co. Record "A" contains pairs 1–25, inclusive. With the exception of pair 28, Record " B," which includes pairs 26–50, is composed of the same pairs contained on Record A, the order of

presentation of the intervals within a combination being reversed. For example, pair 1, which is contained on Record A, consists of the Major Third followed by the Major Seventh, whereas pair 40, which is contained on Record B, consists of the Major Seventh followed by the Major Third. Pairs 13 and 28 are not reversed. In this test the subject is presented with two combinations of two tones each, one combination being better or worse than the other in consonance. A "good combination" is described as one in which the two tones are smooth, and blend, tending to fuse into one. A "bad combination" is just the opposite. This test [1] calls for a sort of composite judgment on *blending, smoothness,* and *fusion,* apart from the feelings of like or dislike, and apart from theory or feeling of musical value. As the various pairs of intervals are played on the phonograph, the subjects make their judgments on the basis of the above criteria, recording *B* if the second combination is better, or *W* if it is worse (*i.e.,* less consonant) than the first.[2]

Eighteen of the subjects who took the foregoing test were classified as possessing "musical training," meaning by this term at least two consecutive years of either vocal or instrumental study. Each student was provided with a record blank which contained spaces for such items of information as name, age, and amount of musical training. After these facts were recorded each student was given a sheet of directions which read as follows:

> You will hear two tones sounded simultaneously, and then, after a brief interval, a second pair. You are to judge which pair is the more consonant. "Consonance" means the tendency of the two tones of a pair to fuse together

[1] For detailed instructions see Seashore's Manual of Instructions and Interpretations, published by the Bureau of Educational Research and Service, University of Iowa, Iowa City.

[2] It will be noted, however, that a slight deviation from Seashore's directions was made in the following experiment, in that the subjects were instructed to base their decisions merely upon two criteria—*fusion* and *blending*—instead of three. However, considerable care was taken in explaining and illustrating relative consonance, and this precaution was regarded as of far more importance than the inclusion or exclusion of a single word such as *smoothness.* This slight change apparently made very little, if any, difference in the decisions of the subjects, since the results secured compared favorably with those obtained by other investigators who adhered strictly to Seashore's directions.

so as to sound like a single tone. You must not judge on the basis of which one you like better or which is more pleasing to you. Disregard your preference and make a detached judgment on the basis of which of the two pairs of tones tends more nearly to fuse or blend into a single unitary sound.

A few illustrative comparisons will be made with tuning-forks before the regular experiment begins. Listen to these carefully and be sure that you understand what you are to listen for before the regular test begins.

Your judgments are to be recorded in terms of the second pair. Thus, if the second pair is better (more consonant) record a B in the appropriate blank on your record sheet; if the second of the two pairs is worse (less consonant), record a W. If you are not sure, *guess;* in case you do guess, encircle this judgment so that the experimenter will know just where you were uncertain.

Do your best to understand and follow these directions absolutely, for the experiment is worthless unless you do. We are trying to study an important problem in the psychology of music and your best cooperation is required for securing reliable results.

The experimenter had a copy of the sheet of directions from which he read while the subjects referred to their individual sheets. After reading the directions the experimenter gave several illustrations of consonance by means of tuning-forks. Special attention was given to the matter of distinguishing between the relative pleasingness and the relative "consonance" of the intervals compared. In this connection several pairs of intervals were presented illustrating the fact that in some instances the more consonant interval is also the more pleasing, whereas in other instances the opposite happens to be true. Upon the completion of the foregoing instructions the Seashore Consonance Record was played in the order A–B. That is, Record "A" was followed by Record "B." After this the records were shifted in such a manner as to convey the impression that a different record was being selected, but in reality the same record was played in reverse (B–A) order. Thus the consonance test was given twice within approximately twenty-five minutes.

2. *The Seashore Consonance Test, with "preference" directions*

Upon the completion of the second consonance test the first set of directions was taken up and another set was given to each subject. This second set of directions required that the subjects use pleasantness and unpleasantness as the basis of comparison. These directions were as follows:

This experiment is similar to the one just finished, except that here you judge on the basis of which pair you like better. Simply listen to the two pairs of tones and if the second combination is more pleasing to you record a P (pleasant); if it is less pleasing record a U (unpleasant). If neither pair is especially pleasing record P or U according as the second pair is more pleasing (P) or less pleasing (U) than the first. Disregard the consonance, if you need to, in any case, and simply let your feeling determine your choice. *Guess,* if you can not be sure about any judgment, but draw a circle around all such uncertain choices.

These directions were read aloud to the subjects while they read from their individual sheets. Upon the completion of the reading of the directions the Seashore Consonance Record was played and the subjects recorded their judgments as before. The record was played twice, in the order A–B, B–A. No attempt was made to conceal the fact that the same record was being used.

The consonance and preference tests were again given on February 15, 1930, to the same group of subjects. The same method of procedure was followed, except that the preference tests were given first, and only one of the consonance tests was given. On April 9, 1930, the consonance test was given again, making a total of four "consonance" and four "preference" tests.

### 3. *Four "consonance tests" of varying difficulty*

On April 16, 1931, another series of experiments was undertaken for the purpose of studying the effects of the presence of very difficult pairs of intervals on the judgment of easy pairs, and also in order to secure additional information concerning the relation of the difficulty of a series of paired-intervals to the consistency of the subjects' judgments. This involved the construction of four different "tests" of varying difficulty. Test 1, shown in table 1, was composed of twenty units—the twenty easiest pairs of intervals of the Seashore Consonance Test, as determined by the percentages of errors shown for these intervals in table 11, column 10; Test 2, shown in table 2, contained twenty-five units, including the same twenty easy pairs used in Test 1, and five difficult pairs which were interspersed at regular intervals; Test 3, shown in table 3, consisted of thirty units, the same twenty easy pairs, with ten difficult pairs; Test 4, shown in

table 4, consisted of twenty units, the ten difficult pairs used in Test 2, and ten other pairs of somewhat less difficulty.

TABLE 1

The 20 "easy" pairs of intervals used in Test 1 of the group of tests varying in difficulty.

| | | |
|---|---|---|
| 1. | Maj. 2nd—Octave | $ga$—$gg'$ |
| 2. | Maj. 3rd—Maj. 7th | $f'a'$—$b^{b}a'$ |
| 3. | Prf. 5th—Octave | $e'^{b}b'^{b}$—$b^{b}b'^{b}$ |
| 4. | Min. 2nd—Prf. 4th | $bc'$—$gc'$ |
| 5. | Maj. 2nd—Min. 2nd | $f'g'$—$f'\sharp g'$ |
| 6. | Octave—Maj. 2nd | $gg'$—$ga$ |
| 7. | Prf. 4th—Min. 2nd | $gc'$—$bc'$ |
| 8. | Prf. 4th—Min. 7th | $d'g'$—$d'c''$ |
| 9. | Min. 6th—Maj. 3rd | $ge'^{b}$—$gb$ |
| 10. | Maj. 2nd—Min. 3rd | $b^{b}c'$—$ac'$ |
| 11. | Maj. 6th—Maj. 7th | $a^{b}f'$—$a^{b}g'$ |
| 12. | Min. 2nd—Octave | $ab^{b}$—$aa'$ |
| 13. | Maj. 2nd—Min. 6th | $ga$—$ge'^{b}$ |
| 14. | Min. 7th—Prf. 4th | $d'c''$—$d'g'$ |
| 15. | Min. 7th—Min. 2nd | $c'b'^{b}$—$c'd'^{b}$ |
| 16. | Aug. 4th—Maj. 7th | $a^{b}d'$—$a^{b}g'$ |
| 17. | Min. 7th—Maj. 2nd | $ag'$—$ab$ |
| 18. | Maj. 7th—Dim. 3rd | $ag'\sharp$—$f'g'\sharp$ |
| 19. | Min. 7th—Maj. 6th | $ba'$—$bg'\sharp$ |
| 20. | Min. 7th—Maj. 7th | $c'b'^{b}$—$c'b'$ |

TABLE 2

Showing the pairs of intervals constituting Test 2 of the tests of varying difficulty. There are 20 easy and 5 difficult pairs. The "difficult" pairs are italicized.

| | | |
|---|---|---|
| 1. | Maj. 2nd—Octave | $ga$—$gg'$ |
| 2. | Maj. 3rd—Maj. 7th | $f'a'$—$b^{b}a'$ |
| 3. | Prf. 5th—Octave | $e'^{b}b'^{b}$—$b^{b}b'^{b}$ |
| 4. | *Min. 6th—Maj. 6th* | $d'b'^{b}$—$d'b'$ |
| 5. | Min. 2nd—Prf. 4th | $bc'$—$gc'$ |
| 6. | Maj. 2nd—Min. 2nd | $f'g'$—$f'\sharp g'$ |
| 7. | Octave—Maj. 2nd | $gg'$—$ga$ |
| 8. | *Min. 2nd—Min. 7th* | $c'd'^{b}$—$c'b'^{b}$ |
| 9. | Prf. 4th—Min. 2nd | $gc'$—$bc'$ |
| 10. | Prf. 4th—Min. 7th | $d'g'$—$d'c''$ |
| 11. | Min. 6th—Maj. 3rd | $ge'^{b}$—$gb$ |
| 12. | *Maj. 2nd—Min. 7th* | $ab$—$ag'$ |
| 13. | Maj. 2nd—Min. 3rd | $b^{b}c'$—$ac'$ |
| 14. | Maj. 6th—Maj. 7th | $a^{b}f'$—$a^{b}g'$ |
| 15. | Min. 2nd—Octave | $ab^{b}$—$aa'$ |
| 16. | *Min. 3rd—Maj. 2nd* | $ac'$—$b^{b}c'$ |
| 17. | Maj. 2nd—Min. 6th | $ga$—$ge'^{b}$ |
| 18. | Min. 7th—Prf. 4th | $d'c''$—$d'g'$ |
| 19. | Min. 7th—Min. 2nd | $c'b'^{b}$—$c'd'^{b}$ |
| 20. | *Dim. 5th—Maj. 3rd* | $e'b'^{b}$—$e'g'\sharp$ |
| 21. | Aug. 4th—Maj. 7th | $a^{b}d'$—$a^{b}g'$ |
| 22. | Min. 7th—Maj. 2nd | $ag'$—$ab$ |
| 23. | Maj. 7th—Min. 3rd | $ag'\sharp$—$f'g'\sharp$ |
| 24. | Min. 7th—Maj. 6th | $ba'bg'\sharp$ |
| 25. | Min. 7th—Maj. 7th | $e'b'^{b}$—$c'b'$ |

## TABLE 3

Showing the pairs of intervals constituting Test 3 of the tests varying in difficulty. There are 20 easy and 10 difficult pairs. The "difficult" pairs are italicized.

| | | |
|---|---|---|
| 1. | *Min. 6th—Maj. 6th* | $d'b'^{b}$—$d'b'$ |
| 2. | Maj. 2nd—Octave | $ga$—$gg'$ |
| 3. | Maj. 3rd—Maj. 7th | $f'a'$—$b^{b}a'$ |
| 4. | *Min. 2nd—Min. 7th* | $c'd'^{b}$—$c'b'^{b}$ |
| 5. | Prf. 5th—Octave | $e'^{b}b'^{b}$—$b^{b}b'^{b}$ |
| 6. | Min. 2nd—Prf. 4th | $bc'$—$gc'$ |
| 7. | *Maj. 2nd—Min. 7th* | $ab$—$ag'$ |
| 8. | Maj. 2nd—Min. 2nd | $f'g'$—$f'\sharp g'$ |
| 9. | Octave—Maj. 2nd | $gg'$—$ga$ |
| 10. | *Min. 3rd—Maj. 2nd* | $ac'$—$b^{b}c'$ |
| 11. | Prf. 4th—Min. 2nd | $gc'$—$bc'$ |
| 12. | Prf. 4th—Min. 7th | $d'g'$—$d'c''$ |
| 13. | *Dim. 5th—Maj. 3rd* | $e'b'^{b}$—$e'g'\sharp$ |
| 14. | Min. 6th—Maj. 3rd | $b^{b}c'$—$ac'$ |
| 15. | Maj. 2nd—Min. 3rd | $b^{b}c'$—$ac'$ |
| 16. | *Maj. 3rd—Dim. 5th* | $e'g'\sharp$—$e'b'^{b}$ |
| 17. | Maj. 6th—Min. 7th | $a^{b}f'$—$a^{b}g'$ |
| 18. | Min. 2nd—Octave | $ab^{b}$—$aa'$ |
| 19. | *Maj. 6th—Maj. 3rd* | $c'a'$—$f'a'$ |
| 20. | Maj. 2nd—Min. 6th | $ga$—$ge'^{b}$ |
| 21. | Min. 7th—Prf. 4th | $d'c''$—$d'g'$ |
| 22. | *Dim. 5th—Min. 6th* | $ae'^{b}$—$af'$ |
| 23. | Min. 7th—Min. 2nd | $c'b'^{b}$—$c'd'^{b}$ |
| 24. | Aug. 4th—Maj. 7th | $a^{b}d'$—$a^{b}g'$ |
| 25. | *Dim. 5th—Min. 6th* | $d'a'^{b}$—$c'a'^{b}$ |
| 26. | Min. 7th—Maj. 2nd | $ag'$—$ab$ |
| 27. | Maj. 7th—Min. 3rd | $ag'\sharp$—$f'g'\sharp$ |
| 28. | *Min. 6th—Min. 3rd* | $g\sharp e'$—$g\sharp b$ |
| 29. | Min. 7th—Maj. 6th | $ba'$—$bg'\sharp$ |
| 30. | Min. 7th—Maj. 7th | $c'b'^{b}$—$c'b'$ |

These four "tests" of varying difficulty were given to a group of 39 students in an introductory psychology class. Each student was provided with a record blank which contained spaces for such items of information as name, age, and amount of musical training. After these facts were recorded the subject was given a sheet of directions which read as follows:

You will hear two tones sounded simultaneously, and then, after a brief interval, a second pair. You are to say which pair is the more consonant. "Consonance" means the tendency of the two tones of a pair to *fuse together* so as to sound like a single tone. You *must not* judge on the basis of which one you like better or which one is more pleasing to you. Disregard your preferences and make a detached judgment on the basis of which of the two pairs of tones tends more nearly to fuse or blend into a unitary sound.

A few illustrative comparisons will be made on the piano before the regular experiment begins. Listen to these carefully and be sure that you understand what you are to listen for, before the regular test begins.

Your judgments are to be recorded in terms of the second pair. Thus, if the second pair is more consonant, record an M in the appropriate blank on

your record sheet; if the second of the two pairs is less consonant, record an L. If you are not sure, *guess.*

Do your best to understand and follow these directions absolutely, for the experiment is worthless unless you do. We are trying to study an important problem in the psychology of music and your best cooperation and attention is required for securing reliable results.

These directions were read aloud by the experimenter while the subjects read from their individual sheets. Upon completion of the reading of the directions several illustrations of intervals differing in consonance were given by the use of the piano. The instrument used was a Steinway Grand piano which had been

TABLE 4

The 20 "difficult" pairs of intervals constituting Test 4 of the tests of varying difficulty.

| | | |
|---|---|---|
| 1. | Min. 3rd—Dim. 5th | e'♭g'♭—c'g'♭ |
| 2. | Min. 3rd—Min. 6th | g♯b—g♯e' |
| 3. | Maj. 7th—Maj. 3rd | b♭a'—g'a' |
| 4. | Min. 3rd—Maj. 7th | f'g'♯—ag'♯ |
| 5. | Dim. 5th—Min. 3rd | c'g'♭—e'♭g'♭ |
| 6. | Maj. 7th—Aug. 4th | a♭g'—a♭d' |
| 7. | Maj. 3rd—Maj. 6th | f'a'—c'a' |
| 8. | Octave—Prf. 5th | b♭b'♭—e'♭b'♭ |
| 9. | Min. 6th—Dim. 5th | af'—ae'♭ |
| 10. | Maj. 6th—Min. 6th | d'b'—d'b'♭ |
| 11. | Min. 6th—Min. 3rd | g♯e'—g♯b |
| 12. | Dim. 5th—Min. 6th | d'a'♭—c'a'♭ |
| 13. | Dim. 5th—Min. 6th | ae'♭—af' |
| 14. | Maj. 6th—Maj. 3rd | c'a'—f'a' |
| 15. | Maj. 3rd—Dim. 5th | e'g'♯—e'b'♭ |
| 16. | Min. 3rd—Maj. 2nd | ac'—b♭c' |
| 17. | Maj. 2nd—Min. 7th | ab—ag' |
| 18. | Min. 2nd—Min. 7th | c'd'♭—c'b'♭ |
| 19. | Min. 6th—Maj. 6th | d'b'♭—d'b' |
| 20. | Dim. 5th—Maj. 3rd | e'b'♭—e'g'♯ |

tuned the day before to the pitch: a'—440 vd. The same precautions were taken with regard to affective-tone and resolution as in the case of the first tests given. In this instance, however, the Perfect Fourth was presented in illustration (without naming the interval) as more consonant than the Major Third. The consecutive pairs of intervals in each of the four tests were then played on the piano, no attempt being made to allow exactly the same amount of time between each interval or each pair of intervals. The attempt was made, however, to present the paired-stimuli fairly rhythmically. An occasional repetition was made upon the request of members of the group. A period of

approximately two minutes elapsed between the completion of one test and the beginning of the next.

The four tests were repeated on the same group of students April 23, 1931, under the same conditions. Just before the tests were repeated, however, favorable comment was made with respect to the results of the first tests. Furthermore, the subjects were cautioned as to the general tendency of individuals to make lower scores on repetitions of such tests, and the hope was expressed that their performance would prove to be an exception to the general rule. After this, a few remarks were made relative to the conditions upon which learning and transfer are dependent, and attention was called to the desirability of putting these laws into operation during the tests which were to follow. After these remarks the same procedure was followed as described for the initial experiment with these tests.

4. *The preferential use of three criteria*

These experiments were carried out for the purpose of discovering whether subjects are more reliable in their judgments when instructed as to the preferential use (*infra,* p. 38) of the criteria *blending, smoothness,* and *purity.*

In order to carry out this study the same pairs (with the exception of pairs 19 and 47[1]) of intervals were used, and in the same order, as that of the Seashore Consonance Test Record. The test was given for the first time on April 17, 1931, at 8 A.M. to a group of 39 students in an introductory psychology class. Before giving the regular directions a few introductory remarks concerning the general nature of the experiment were made, and the necessity for the maintenance of a uniform degree of attention was emphasized.

Each subject was provided with a record blank similar to those used in previous tests. After recording the desired supplementary information each subject was given a sheet of directions which read as follows:

> The present test was devised for the purpose of studying your ability to discriminate the relative consonance of pairs of tones. Consonance means the tendency of two tones to blend together so as to sound like a single tone.

[1] The substitutions were: pair 19: a g′♯—f′g′♯; pair 47: f′g′♯—a g′♯.

You will hear two tones (constituting an "interval") sounded simultaneously, and then, after a brief period, a second pair. You are to judge which pair is the more consonant.

In general, the more consonant intervals BLEND better, are SMOOTHER and PURER than less consonant intervals.

(1) BLENDING—a seeming to agree, to belong together.

(2) SMOOTHNESS—a relative freedom from beats.

(3) PURITY—thinness of tone, absence of richness.

You will hear two tones sounded simultaneously, and then, after a brief period, a second pair. You are to judge which pair is the more consonant.

Give your decision on BLENDING alone if the degree of blending is perceptibly different; if not, make the decision on SMOOTHNESS; and if there is no difference in either blending or smoothness, base your decision on PURITY.

Keeping these criteria in mind, if the second pair is *more* consonant than the first, record a capital *M* in the appropriate square. If the second pair is *less* consonant than the first (judgment is always recorded in terms of the second pair), record an *L*.

In the upper left-hand corner of each square indicate which criterion your decision was based on; thus, if the factor of BLENDING was the basis of your judgment, indicate this fact by placing a small *b* in the upper left-hand corner of that particular square; if it was SMOOTHNESS, place a small *s;* if it was PURITY, record a *p*.

In case you can not decide which pair is the more consonant, GUESS—leave no blanks.

The experimenter read the directions to the group from a similar sheet while the subjects read from their individual sheets. After the reading of the directions the experimenter illustrated the meaning of the three criteria by playing certain pairs of intervals on the piano. The same precautions were taken with respect to affective-tone as in preceding tests. The pairs of intervals constituting this series were then played on the piano in approximately the same manner as in the preceding series, although the subjects were allowed slightly more time in which to record their judgments and the supplementary notations concerning the criteria employed.

This experiment was repeated April 24, 1931, at 8 A.M. on the same group of students, under the same conditions. Before the experiment began a few introductory remarks were made in which the subjects were told of the general tendency to make lower scores upon repetitions of the test, and of the consequent necessity of maintaining a uniformly high degree of attention throughout the test, in order to avoid such a result. Upon the completion of the directions the paired-intervals were played in

the same manner as at the first sitting. Thus the series of intervals was presented to this group twice, in the same order and under the same conditions.

### 5. *The use of each of two criteria in separate series*

This experiment, which involved the use of two tests (called here the "Preliminary test" and the "Single Criterion test") was made chiefly for the purpose of determining whether subjects are more consistent in their judgments when the pairs of intervals are so arranged that a single criterion may be used at a time. The main test, termed "Single Criterion test," consisted of fifty units—the fifty paired-intervals of the Seashore Consonance Test, arranged in two parts. Part I, shown in table 5, consisted of twenty-nine pairs which, according to the experimenter's judgment, could best be judged on the basis of *smoothness*.

TABLE 5

*Showing pairs of intervals for Part I of Single Criterion Test*

| | | |
|---|---|---|
| 1. | Maj. 3rd—Maj. 7th | $f'a'$—$b^{b}a'$ |
| 2. | Aug. 4th—Maj. 7th | $a^{b}d'$—$a^{b}g'$ |
| 3. | Min. 7th—Maj. 2nd | $ag'$—$ab$ |
| 4. | Maj. 3rd—Maj. 6th | $f'a'$—$c'a'$ |
| 5. | Min. 7th—Maj. 7th | $c'b'^{b}$—$c'b'$ |
| 6. | Min. 2nd—Maj. 2nd | $f'\sharp g'$—$f'g'$ |
| 7. | Maj. 6th—Maj. 7th | $a^{b}f'$—$a^{b}g'$ |
| 8. | Min. 2nd—Octave | $ab^{b}$—$aa'$ |
| 9. | Min. 7th—Min. 2nd | $c'b'^{b}$—$c'd'^{b}$ |
| 10. | Prf. 4th—Min. 2nd | $gc'$—$bc'$ |
| 11. | Maj. 2nd—Min. 6th | $ga$—$ge'^{b}$ |
| 12. | Maj. 7th—Min. 3rd | $ag'\sharp$—$f'g'\sharp$ |
| 13. | Maj. 6th—Min. 7th | $bg'\sharp$—$ba'$ |
| 14. | Prf. 4th—Maj. 3rd | $c'f'$—$c'e'$ |
| 15. | Octave—Maj. 2nd | $gg'$—$ga$ |
| 16. | Min. 2nd—Prf. 4th | $bc'$—$gc'$ |
| 17. | Min. 2nd—Min. 7th | $c'd'^{b}$—$c'b'^{b}$ |
| 18. | Maj. 7th—Maj. 6th | $a^{b}g'$—$a^{b}f'$ |
| 19. | Maj. 2nd—Min. 2nd | $f'g'$—$f'\sharp g'$ |
| 20. | Maj. 7th—Min. 7th | $c'b'$—$c'b'^{b}$ |
| 21. | Maj. 6th—Maj. 3rd | $c'a'$—$f'a'$ |
| 22. | Maj. 2nd—Min. 7th | $ab$—$ag'$ |
| 23. | Maj. 7th—Aug. 4th | $a^{b}g'$—$a^{b}d'$ |
| 24. | Maj. 7th—Maj. 3rd | $b^{b}a'$—$f'a'$ |
| 25. | Maj. 2nd—Octave | $ga$—$gg'$ |
| 26. | Maj. 3rd—Prf. 4th | $c'e'$—$c'f'$ |
| 27. | Min. 7th—Maj. 6th | $ba'$—$bg'\sharp$ |
| 28. | Min. 3rd—Maj. 7th | $f'g'\sharp$—$ag'\sharp$ |
| 29. | Min. 6th—Maj. 2nd | $ge'^{b}$—$ga$ |

Part II, shown in table 6, was composed of twenty-one pairs which could best be judged on the basis of *blending*.

This Single Criterion test was given for the first time on January 22, 1932, at 8 A.M. to a group of 32 students in an introductory psychology class. Each subject was provided with a record blank for the usual items of information.

The " Preliminary test " was given solely for the purpose of affording a check upon the comparative reliability of the Single

TABLE 6

*Showing pairs of intervals for Part II of Single Criterion Test*

| | |
|---|---|
| 1. Prf. 5th—Octave | $e'^{b}b'^{b}$—$b^{b}b'^{b}$ |
| 2. Min. 3rd—Min. 6th | g♯b—g♯e' |
| 3. Dim. 5th—Min. 6th | $d'a'^{b}$—$c'a'^{b}$ |
| 4. Maj. 3rd—Min. 6th | gb—$ge'^{b}$ |
| 5. Min. 3rd—Maj. 2nd | ac'—$b^{b}c'$ |
| 6. Dim. 5th—Maj. 3rd | $e'b'^{b}$—e'g'♯ |
| 7. Min. 7th—Prf. 4th | d'c''—d'g' |
| 8. Min. 3rd—Dim. 5th | $e'^{b}g'^{b}$—$c'g'^{b}$ |
| 9. Dim. 5th—Min. 6th | $ae'^{b}$—af' |
| 10. Min. 6th—Maj. 6th | $d'b'^{b}$—d'b' |
| 11. Octave—Maj. 3rd | aa'—ac'♯ |
| 12. Maj. 2nd—Min. 3rd | $b^{b}c'$—ac' |
| 13. Min. 6th—Maj. 3rd | $ge'^{b}$—gb |
| 14. Min. 6th—Dim. 5th | $c'a'^{b}$—$d'a'^{b}$ |
| 15. Min. 6th—Min. 3rd | g'♯e'—g♯b |
| 16. Octave—Prf. 5th | $b^{b}b'^{b}$—$e'^{b}b'^{b}$ |
| 17. Maj. 6th—Min. 6th | d'b'—$d'b'^{b}$ |
| 18. Min. 6th—Dim. 5th | af'—$ae'^{b}$ |
| 19. Dim. 5th—Min. 3rd | $c'g'^{b}$—$e'^{b}g'^{b}$ |
| 20. Prf. 4th—Min. 7th | d'g'—d'c'' |
| 21. Maj. 3rd—Dim. 5th | e'g'♯—e'b'♯ |

Criterion Test, *i.e.*, as a check upon the consistency of consonance judgments when the pairs of intervals are so arranged that a single criterion can be used at a time. The same pairs of intervals were used in this test as in the main test, and the order of presentation was identical with that shown for the Seashore test in table 11. The directions for this test were approximately the same as those for experiment 4 (*supra*, p. 38) except that the subjects were instructed to base their decisions on *smoothness*, if the degree of smoothness was perceptibly different, and if not, to make their decisions on the basis of *blending*. Upon the completion of this preliminary test each subject was given a list of directions which read as follows:

### Single Criterion Test

The present test was devised for the purpose of studying your ability to discriminate the relative consonance of pairs of tones. CONSONANCE means the tendency of two tones to blend together so as to sound like a single tone.

You will hear two tones (constituting an "interval") sounded simultaneously, and then, after a brief period, a second pair. You are to judge which pair is the more consonant.

In general, the more consonant intervals are SMOOTHER and BLEND better than less consonant intervals.

(1) SMOOTHNESS—a relative freedom from beats, absence of roughness.

(2) BLENDING—a seeming to agree, to belong together, absence of strangeness.

Part I

In this series of pairs you are to base your decisions on SMOOTHNESS only. Thus if the second pair is smoother than the first, it is more consonant than the first and you will record an *M* in the appropriate square. If the second pair (judgment is always recorded in terms of second pair) is less smooth than the first, it is less consonant and you will record an *L* in the proper square.

Part II.

In this series of pairs you are to use the criterion of BLENDING as the SOLE basis for your decisions. You will hear two tones (an interval) sounded simultaneously, and then, after a brief period, a second pair. If the second pair BLENDS better than the first, it is more consonant than the first and you will record an *M* in the appropriate square. If it does not blend as well as the first you will record an *L* in the square.

On January 27, 1932, the Preliminary test and the Single Criterion test were repeated with the same group and under the same conditions, with the exception that the latter test was presented first.

# CHAPTER IV

## General Analysis of the Seashore Consonance Test "Scores"

Before analyzing the results of the several experiments on the effects of paired-interval difficulty, affective-tone, and various criteria upon consonance judgments, the gross "scores" made by the subjects on the several applications of the Seashore Consonance Test will be considered. As previously mentioned, the Seashore Consonance Record was played four times, twice under each of the following two sets of instructions: (1) that comparisons be made on the basis of consonance value, and (2) that judgments be made on the basis of preference, with no suggestion as to analysis of the bases of preferences. The average efficiency, the effects of repetition upon efficiency, and the "reliability" of the judgments will be shown. Following the study of the group as a whole, a division of the subjects into musically trained and untrained will be made, and the two sub-groups compared as to "efficiency" and "reliability."

### 1. *The general efficiency of the subjects on the "consonance" and "preference" tests*

The means, standard deviations, and the reliability of the differences between the several means of the "scores" (per cent of correct judgments) secured with the use of the Seashore Consonance Record are shown in table 7. The mean score for the first "consonance" series is 67 per cent of correct judgments, whereas that for the first "preference" series is 68 per cent correct. These results agree well with those of other investigators. In a group of 200 university students Weaver found a mean of 69.17 per cent correct, with a standard deviation of 4.43 (**30**, p. 170). The mean for Larson's group of 132 cases was 67.41, with a standard deviation of 4.09 (**14**, p. 58). It would seem that the average "score" of the present group is

TABLE 7

***Showing the means, in terms of per cent correct, the standard deviations, and the reliability of the differences between certain averages for 36 subjects***

| | Consonance | | | | Preference | | | |
|---|---|---|---|---|---|---|---|---|
| Tests | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Mean | 67 | 63 | 63 | 62 | 68 | 66 | 65 | 64 |
| S.D. | 9.20 | 8.05 | 8.20 | 8.25 | 8.00 | 8.30 | 8.60 | 8.95 |
| Tests Compared | 1–2 | 2–3 | 3–4 | 1–4 | 1–2 | 2–3 | 3–4 | 1–4 |
| Diff. / Sigma Diff. | 1.97 | 0 | .52 | 1.83 | 1.05 | .50 | .48 | 2.01 |

representative of the general efficiency of unselected college students in making the comparative judgments of consonance under these experimental conditions. The standard deviation of our group is, however, about twice as great as that in the two studies cited. This greater variability of scores may be due to the smaller number of subjects used, since an occasional high or low score would affect the standard deviation to a greater degree in the case of a small group.

There is a fairly progressive decrease in the means upon the repetition of both the consonance and the preference series, and although the differences are not great for successive tests, they are sufficiently large between the initial and the final tests to be indicative of a real decrease in efficiency in judging relative consonance. In the case of the four consonance tests the difference in average score between the first and the last is 5 per cent, while the average difference between the first and last preference tests is 4 per cent. The reliabilities of these differences, in terms of the critical index Diff./Sigma Diff., are 1.83 and 2.01, respectively. Although the above differences are not statistically reliable they are sufficiently large to make it highly probable (96 chances in 100 in the case of the consonance tests, and 98 in 100 for the preference tests) that the true difference in each case is greater than zero. These findings agree in general with results secured by other investigators. In a group of 35 subjects Heinlein (**9**, p. 419) found a general increase in the number of errors upon repetition of the Seashore Consonance Test. Larson (**14**, p. 58) on the other hand, reports that she found no

general tendency to score either higher or lower on the repetition of this test. In general, however, Heinlein's results seem to be the more typical, and, in view of the nature of the test, also the more logical. If we take into consideration the fact that there is nothing intrinsically interesting in this test to the average subject, it is perhaps to be expected that a decrease in efficiency would result from repetition. It has already been suggested that many of these comparisons are difficult and require the maintenance of an alert, highly "discriminative" attitude. This is difficult to secure, since it depends upon unusual motivation and perhaps upon a type of training not present in subjects of this sort.

### 2. *Comparative efficiencies of musically trained and untrained subjects*

Some indication of the validity of the Seashore Consonance Test as a measure of this aspect of tonal discriminability may perhaps be given by a comparison of the records of musically trained and untrained subjects. Heinlein claims that musically trained subjects tend to score lower on the test than the untrained and uses this supposed fact as an argument against the significance of the test scores. A comparison of the mean scores of the trained and of the untrained subjects in our group has been made as a check on this conclusion. The results, given in table 8, show that the 18 subjects with musical training are slightly more efficient than the 18 untrained subjects. In the case of the "consonance" judgments, their respective averages are 65.2 and

TABLE 8

*Showing means, in terms of per cent correct, and standard deviations of 18 trained and 18 untrained subjects on the four consonance and the four preference tests*

| | Consonance | | | | Preference | | | |
|---|---|---|---|---|---|---|---|---|
| | Untrained | | Trained | | Untrained | | Trained | |
| Tests | Mean | S.D. | Mean | S.D. | Mean | S.D. | Mean | S.D. |
| 1 | 65 | 8.20 | 69 | 9.70 | 66 | 7.07 | 72 | 7.65 |
| 2 | 62 | 8.80 | 65 | 6.90 | 64 | 5.95 | 69 | 9.50 |
| 3 | 61 | 6.25 | 65 | 9.30 | 63 | 8.05 | 67 | 8.66 |
| 4 | 61 | 7.20 | 62 | 9.15 | 62 | 7.75 | 66 | 9.50 |
| Average | 62.5 | | 65.2 | | 63.7 | | 68.5 | |

62.5 per cent correct. This disparity is greater for the "preference" judgments, the untrained having an average mean of 63.7 per cent correct as against 68.5 per cent for the trained group. These differences are not statistically reliable, the critical index Diff./Sigma Diff. being 1.34, and 2.44 for the "consonance" and "preference" differences, respectively. Other investigators have also found musically trained subjects to be superior to those without such training. In a recent study Brown found the mean for the untrained to be 63.74 per cent, whereas the mean for those with training was 69.52 per cent (**1**, p. 49). The difference in favor of the trained subjects is 5.78, the probable error of the difference being 1.97. Larson (**14**, p. 62) found that subjects with musical training averaged much higher than those without such training. For a group of 35 musically trained subjects Larson found a mean of 75.88, whereas for a group of untrained subjects the mean was 63.49. In order to determine whether the size of the group had any bearing on the results Larson also compared 35 musically trained subjects with 35 untrained subjects, giving the tests individually. Apparently, the size of the group is unimportant, since under these conditions the trained subjects had an average of 78.75, while those without training had an average of only 60.79. On the other hand, Heinlein (**8**, p. 419) found a tendency for subjects with training to make lower scores than untrained subjects. In general, however, experimental results seem to indicate that individuals with musical training are more efficient than those without such training, although the superiority is not great.

### 3. *The general reliability of the "consonance" and "preference" tests*

Successive applications of the Seashore Consonance Test have been found to yield somewhat inconsistent results, as indicated by coëfficients of correlation between the scores made by subjects on such repeated tests. The "reliability" coefficients secured by previous investigators have varied from .35 to .68. The following "reliability coefficients" indicate the relatively low degree of "self-correlation" which has been obtained for this test:

Gaw (3, p. 305), .49±.08; Lanier (13, p. 93), .54±.05; Peterson (20, p. 32), .68±.04 for 89 white students, and .52±.03 for 197 negroes; Ruch and Stoddard (23, p. 195), .35±.06; Heinlein (3, p. 305), .62±.07, for one group, and .48±.09, for a second group. In a somewhat recent study by Larson (14, p. 57), the attempt was made to determine the reliability of the consonance test when Seashore's directions are presumably more carefully followed. Retests of two groups showed reliabilities of .63±.02, and .65±.02. As already pointed out by Farnsworth (3, p. 306) and by Heinlein (9, p. 532), these coefficients are too low for the degree of reliability usually required for a "test."

The "reliability coefficients" for the present study are shown in table 9, and they are in fairly good agreement with those just cited. The intercorrelations among the four applications of the consonance test vary from .42 to .65, the average being .52. The "preference" judgments are more reliable, the intercorrelations among the several repetitions varying from .46 to .68, with an average $R$ (reliability coefficient) of .60.

This low reliability means, of course, that the conditions determining the performances of the subjects are not uniform for each individual at a given sitting, and that such conditions vary in successive sittings in a differential fashion. The logical conclusion from such considerations is that a "test" of consonance discriminability does not exist and would be difficult if not impossible of realization. A "test" requires that the performance in question be defined as a function of definite conditions so that

TABLE 9

*Showing the reliability coefficient for the four applications of the "consonance" and the "preference" tests based on the gross scores, in terms of per cent correct, of 36 subjects.*

| Series | Consonance | | Preference | |
|---|---|---|---|---|
| Correlated | $R$ | P.E. | $R$ | P.E. |
| 1–2 | .53 | .08 | .66 | .06 |
| 1–3 | .56 | .08 | .61 | .07 |
| 1–4 | .48 | .08 | .46 | .09 |
| 2–3 | .42 | .09 | .68 | .06 |
| 2–4 | .48 | .08 | .65 | .06 |
| 3–4 | .65 | .06 | .59 | .08 |
| Average | .52 | | .60 | |

it can be assumed that any subsequent performance under similar conditions would be the equivalent of the earlier performance. There are no perfect psychological " tests," but there are better ones than the present consonance test. The index of the consistency of the operation of the causal factors involved (" reliability coefficient ") is too low here to justify the hope of using this type of " score " for the purposes to which such " test " scores ordinarily are put. That is to say, correlations between the successive " measurements " of consonance discrimination are so low as to make any individual prediction from one application practically a matter of chance. It will be shown below that neither such gross " scores " as these nor the " reliability " coefficients based upon them serve as accurate scientific descriptions of the specific behavioral operations involved.

4. *Comparative reliabilities of musically trained and untrained subjects*

Since our group was composed of 18 musically trained subjects and 18 subjects without musical training, it seemed desirable to ascertain the effects of training upon consistency in consonance judgments. Accordingly, " reliability coefficients " have been computed for these two sub-groups and they are presented in table 10.[1] In the case of the consonance judgments the trained

TABLE 10

*Reliability coefficients for 18 trained and 18 untrained subjects on the four consonance and four preference series*

| Series Correlated | Consonance | | | | Preference | | | |
|---|---|---|---|---|---|---|---|---|
| | Untrained | | Trained | | Untrained | | Trained | |
| | *R* | P.E. | *R* | P.E. | *R* | P.E. | *R* | P.E. |
| 1–2 | .61 | .10 | .51 | .12 | .57 | .11 | .65 | .09 |
| 1–3 | .41 | .13 | .63 | .09 | .51 | .12 | .64 | .09 |
| 1–4 | .25 | .15 | .61 | .10 | .54 | .11 | .34 | .14 |
| 2–3 | .48 | .12 | .31 | .21 | .43 | .13 | .82 | .05 |
| 2–4 | .44 | .13 | .54 | .11 | .65 | .09 | .63 | .09 |
| 3–4 | .61 | .10 | .66 | .09 | .45 | .13 | .67 | .09 |
| Average | .466 | | .543 | | .525 | | .625 | |

[1] As will be shown later (*infra*, pp. 58–59), a reliability coefficient based on gross scores in terms of per cent correct is not a satisfactory index of subjects' consistency in judging relative consonance, and the *R*s contained in the present analysis are presented with this qualification in mind. However, they agree in general with those obtained by Larson and other writers for musically trained and untrained subjects.

subjects, with an average *R* of .543, are more consistent than the untrained who have an average *R* of .466. This superiority is greater in the case of the preference series, the trained subjects having an *R* of .625, as against .525 for the untrained. These results agree, in general, with those of Larson (**14**, p. 59) who found that subjects with musical training are more consistent than untrained subjects. On the other hand, Heinlein (**8**, p. 433) states, on the basis of rather meager data, that musically trained subjects are less reliable than those without such training. He points out that " from the very nature and structure of the test material, there is every reason to expect negative results from the talented group." Regardless of the more or less *a priori* speculation which Heinlein emphasizes, experimental results indicate that musically trained subjects are somewhat more consistent in such tonal judgments than the untrained. It should be noted here that the criteria of " training " differ with different investigators and that this fact may account for discrepancies in their comparisons of trained with untrained subjects.

# CHAPTER V

## The Effects of Paired-Interval Difficulty on Consonance Judgments

The present chapter is devoted to a study of the effects of the difficulty of the judgments required of the subjects upon the "accuracy" (with respect to a more or less arbitrary standard of accuracy) and consistency of such judgments as may be required in a series of paired-interval comparisons. As previously indicated, the problem of paired-interval difficulty has received the consideration of only a few investigations in this field. In devising a test for consonance discrimination, Malmberg (17) attempted to make allowance for the fact that certain pairs of intervals are more difficult to judge on the basis of relative consonance than others. Gaw (4) continued the testing program initiated by Malmberg, revising the latter's sixty-six unit test by the elimination of certain pairs of intervals which she regarded as either too difficult or too easy. However, the problem of paired-interval difficulty is of much wider significance for consonance discrimination than the slight consideration given it by the above writers would seem to indicate. Inasmuch as the character of any response is determined partly by the nature of the stimulus, the accuracy and consistency of consonance judgments should be conditioned to some extent by the difficulty of the paired-intervals used to secure the judgments. Obviously, the establishment of such a relationship would have an important bearing upon the validity of any consonance test such as Seashore's, which evidently was devised without due regard for the effect of difficult pairs of intervals upon consistency of judgment. However, the seriousness of this oversight will become increasingly manifest in the following discussions.

It should be noted here that the phrase "paired-interval difficulty" is equivocal, since two distinct types of factors produce the "errors" on the basis of which "difficulty" is estimated. In one sense, a difficult comparison would involve two intervals

which were quite similar and hence difficult to discriminate with respect to relative degree of consonance. Such judgments might be of a purely fortuitous nature. Another type of "difficulty" is involved when a consistent, but "easily made," judgment is rendered in favor of the interval considered to be the more dissonant. It will be shown later that the pleasant affective-tone induced by "resolution" may cause the subject to be biased in favor of an interval which is decidedly the more dissonant of a pair, in terms of the usual criteria. Thus the percentage of errors in the case of such a pair would be high and on this basis the pair would be called "difficult." Yet the judgment is "easy" for the subject in the sense that it is made readily and with confidence. Thus "difficulty" may mean actual confusion in comparing intervals which are very similar in consonance value, or it may refer to the inability of the subject to disregard irrelevant factors which make his judgment "easy," yet "erroneous." No attempt is made here to specify in detail the paired-intervals falling in these respective categories, although in Chapter VI, the more outstanding examples of the second type of "difficulty" will be analyzed. "Paired-interval difficulty" is estimated here in terms of the per cent of erroneous judgments made for the combinations, irrespective of cause.

The relative difficulty of the fifty pairs of intervals in the Seashore test was determined by tabulating the number of errors (according to Seashore's standards) made by the 36 subjects upon each combination in each application of the "test." These data permit a study of the consistency of the error-frequency values in the several applications of the tests and provide material for later analyses. Second, on the basis of the error-frequency values sub-tests of four grades of difficulty are created by artificial subdivision of the subjects' Seashore test records, and the "reliabilities" of these sub-tests are studied. Third, a new series of tests is constructed, of four grades of difficulty, and applied to a different group of subjects in order to compare the consistency of judgment under such conditions with that in the Seashore test, where easy and difficult pairs are indiscriminately mixed.

### 1. *The relative difficulty of the combinations of intervals in the Seashore Consonance Test*

The numbers and percentages of errors for each combination in the Seashore test, on all applications, are shown in table 11; the "error-frequencies" for both "consonance" and "preference" tests are given. This table contains data of a nature not hitherto available in the literature, and which permit several types of analysis made for the first time in this study. The first column contains the numbers of the combinations of intervals in consecutive order, while the second column gives the intervals constituting these combinations. The numbers in parentheses in column 1 indicate the numbers of these combinations which have the same intervals but with the order of presentation reversed. Thus, combination 1 consists of a Major Third followed by a Major Seventh, while combination 40 consists of a Major Seventh followed by a Major Third. It will be shown later that the order of presentation is often a determining factor in the error-frequency of a given combination. Combinations 13 and 28 are presented only once in the test. The meaning of the figures in table 11 should be clear, in view of the description of the several applications of the Seashore test in Chapter III. The test was given eight times to the same 36 subjects, four times with directions to compare the intervals solely as to relative degree of *consonance,* and four times with instructions to judge on the basis of *preference* solely. The figures in table 11 are the numbers and per cents of errors for each of the fifty comparisons in the eight applications of the Seashore test, as made under the conditions mentioned above. For example, column 3 (the first regular column in the table) shows that a total of two errors was made by the 36 subjects on combination 1 for the first application of the Seashore test with "consonance" directions; the fourth column shows an error-frequency of two for the same pair for the second application of the "consonance" test; the total number of errors for the first and second consonance series, 4, is shown in the fifth column; and the corresponding per cent of errors, 6, is shown in column 6. Columns 7, 8, 9, and 10 show corresponding results for the third and fourth presentations of

TABLE 11

*Showing the error-frequencies for each pair of intervals in the Seashore test in the four "consonance" and the four "preference" series. The italicized combination is considered by Seashore to be the more consonant. The values are based on the records of 36 subjects.*

| Pair no. | Combination | Consonance Series 1 No. of errors | 2 No. of errors | 1 and 2 Combined Total | 1 and 2 Combined Pct. | 3 No. of errors | 4 No. of errors | 3 and 4 combined Total | 3 and 4 combined Pct. | Preference Series 1 No. of errors | 2 No. of errors | 1 and 2 combined Total | 1 and 2 combined Pct. | 3 No. of errors | 4 No. of errors | 3 and 4 combined Total | 3 and 4 combined Pct. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 (40) | *Maj. 3rd*—Maj. 7th | 2 | 2 | 4 | 6 | 3 | 2 | 5 | 7 | 0 | 0 | 0 | 0 | 1 | 3 | 4 | 6 |
| 2 (39) | *Aug. 4th*—Maj. 7th | 4 | 9 | 13 | 18 | 11 | 11 | 22 | 30 | 4 | 8 | 12 | 17 | 4 | 10 | 14 | 19 |
| 3 (38) | Prf. 5th—*Octave* | 6 | 6 | 12 | 17 | 5 | 3 | 8 | 11 | 7 | 6 | 13 | 18 | 10 | 13 | 23 | 32 |
| 4 (37) | Min. 3rd—*Min. 6th* | 11 | 12 | 23 | 32 | 21 | 14 | 35 | 49 | 5 | 7 | 12 | 17 | 7 | 5 | 12 | 17 |
| 5 (36) | Dim. 5th—*Min. 6th* | 22 | 23 | 45 | 62 | 17 | 24 | 41 | 57 | 17 | 17 | 34 | 47 | 20 | 14 | 34 | 47 |
| 6 (35) | *Min. 7th*—Maj. 2nd | 14 | 18 | 32 | 44 | 10 | 12 | 22 | 30 | 10 | 8 | 18 | 25 | 14 | 9 | 23 | 32 |
| 7 (34) | Maj. 3rd—*Maj. 6th* | 17 | 13 | 30 | 42 | 20 | 17 | 37 | 51 | 11 | 11 | 22 | 30 | 12 | 13 | 25 | 36 |
| 8 (33) | *Maj. 3rd*—Min. 6th | 15 | 14 | 29 | 40 | 17 | 11 | 28 | 39 | 25 | 23 | 48 | 67 | 19 | 27 | 46 | 64 |
| 9 (32) | *Min. 7th*—Maj. 7th | 6 | 1 | 7 | 10 | 11 | 13 | 24 | 33 | 3 | 3 | 6 | 8 | 3 | 6 | 9 | 12 |
| 10 (31) | Min. 2nd—*Maj. 2nd* | 8 | 11 | 19 | 26 | 12 | 13 | 25 | 36 | 5 | 7 | 12 | 17 | 3 | 6 | 9 | 12 |
| 11 (30) | *Min. 3rd*—Maj. 2nd | 16 | 21 | 37 | 51 | 23 | 21 | 44 | 61 | 25 | 22 | 47 | 65 | 20 | 24 | 44 | 61 |
| 12 (29) | *Maj. 6th*—Maj. 7th | 6 | 6 | 12 | 17 | 8 | 9 | 17 | 24 | 3 | 8 | 11 | 15 | 4 | 11 | 15 | 21 |
| 13 | Min. 2nd—*Octave* | 7 | 13 | 20 | 28 | 6 | 12 | 18 | 25 | 10 | 7 | 17 | 24 | 15 | 16 | 31 | 43 |
| 14 (27) | *Min. 7th*—Min. 2nd | 6 | 8 | 14 | 19 | 9 | 11 | 20 | 28 | 0 | 3 | 3 | 4 | 1 | 3 | 4 | 6 |
| 15 (26) | *Prf. 4th*—Min. 2nd | 2 | 7 | 9 | 12 | 6 | 7 | 13 | 18 | 2 | 4 | 6 | 8 | 1 | 7 | 8 | 11 |
| 16 (50) | Dim. 5th—*Maj. 3rd* | 23 | 25 | 48 | 67 | 19 | 24 | 43 | 60 | 17 | 21 | 38 | 53 | 22 | 22 | 44 | 61 |
| 17 (49) | Maj. 2nd—*Min. 6th* | 5 | 4 | 9 | 12 | 11 | 7 | 18 | 25 | 7 | 5 | 12 | 17 | 2 | 1 | 3 | 4 |
| 18 (48) | Min. 7th—*Prf. 4th* | 5 | 9 | 14 | 19 | 12 | 7 | 19 | 26 | 6 | 9 | 15 | 21 | 3 | 5 | 8 | 11 |
| 19 (47) | *Maj. 7th*—Min. 2nd | 9 | 11 | 20 | 28 | 12 | 10 | 22 | 30 | 2 | 2 | 4 | 6 | 1 | 1 | 2 | 3 |
| 20 (46) | *Min. 3rd*—Dim. 5th | 11 | 15 | 26 | 36 | 18 | 14 | 32 | 44 | 13 | 16 | 29 | 40 | 20 | 22 | 42 | 58 |
| 21 (45) | Dim. 5th—*Min. 6th* | 17 | 22 | 39 | 54 | 19 | 22 | 41 | 57 | 10 | 17 | 27 | 37 | 13 | 13 | 26 | 36 |
| 22 (44) | *Maj. 6th*—Min. 7th | 13 | 5 | 18 | 25 | 16 | 12 | 28 | 39 | 14 | 10 | 24 | 33 | 8 | 11 | 19 | 26 |
| 23 (43) | Prf. 4th—*Maj. 3rd* | 15 | 14 | 29 | 40 | 12 | 14 | 26 | 36 | 11 | 11 | 22 | 30 | 11 | 4 | 15 | 21 |
| 24 (42) | *Octave*—Maj. 2nd | 1 | 4 | 5 | 7 | 6 | 4 | 10 | 14 | 2 | 5 | 7 | 10 | 6 | 5 | 11 | 15 |
| 25 (41) | Min. 6th—*Maj. 6th* | 23 | 29 | 52 | 72 | 27 | 28 | 55 | 76 | 28 | 25 | 53 | 74 | 22 | 28 | 50 | 69 |

TABLE 11—*Continued*

*Showing the error-frequencies for each pair of intervals in the Seashore test in the four "consonance" and the four "preference" series. The italicized combination is considered by Seashore to be the more consonant. The values are based on the records of 36 subjects.*

| | | Consonance Series | | | | | | | | Preference Series | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 1 and 2 | | 3 | 4 | 3 and 4 | | 1 | 2 | 1 and 2 | | 3 | 4 | 3 and 4 | |
| | | No. of | No. of | Combined | | No. of | No. of | combined | | No. of | No. of | combined | | No. of | No. of | combined | |
| Pair no. | Combination | errors | errors | Total | Pct. | errors | errors | Total | Pct. | errors | errors | Total | Pct. | errors | errors | Total | Pct. |
| 26 (15) | Min. 2nd—*Prf. 4th* | 6 | 4 | 10 | 14 | 3 | 6 | 9 | 12 | 4 | 1 | 5 | 7 | 4 | 7 | 11 | 15 |
| 27 (14) | Min. 2nd—*Min. 7th* | 26 | 24 | 50 | 69 | 18 | 30 | 48 | 67 | 29 | 29 | 58 | 80 | 29 | 31 | 60 | 83 |
| 28 | *Octave*—Maj. 3rd | 18 | 25 | 43 | 60 | 17 | 12 | 29 | 40 | 27 | 26 | 53 | 74 | 31 | 28 | 59 | 82 |
| 29 (12) | Maj. 7th—*Maj. 6th* | 4 | 7 | 11 | 15 | 14 | 12 | 26 | 36 | 4 | 4 | 8 | 11 | 6 | 3 | 9 | 12 |
| 30 (11) | Maj. 2nd—*Min. 3rd* | 5 | 7 | 12 | 17 | 7 | 9 | 16 | 22 | 5 | 6 | 11 | 15 | 7 | 8 | 15 | 21 |
| 31 (10) | *Maj. 2nd*—Min. 2nd | 5 | 3 | 8 | 11 | 5 | 4 | 9 | 12 | 0 | 1 | 1 | 1 | 2 | 0 | 2 | 3 |
| 32 (9) | Maj. 7th—*Min. 7th* | 12 | 7 | 19 | 26 | 18 | 14 | 32 | 44 | 4 | 9 | 13 | 18 | 8 | 4 | 12 | 17 |
| 33 (8) | Min. 6th—*Maj. 3rd* | 11 | 11 | 22 | 30 | 8 | 7 | 15 | 21 | 6 | 4 | 10 | 14 | 5 | 7 | 12 | 17 |
| 34 (7) | *Maj. 6th*—Maj. 3rd | 22 | 19 | 41 | 57 | 22 | 19 | 41 | 57 | 25 | 22 | 47 | 65 | 25 | 22 | 47 | 65 |
| 35 (6) | Maj. 2nd—*Min. 7th* | 26 | 28 | 54 | 75 | 24 | 22 | 46 | 64 | 27 | 28 | 55 | 76 | 25 | 23 | 48 | 67 |
| 36 (5) | *Min. 6th*—Dim. 5th | 20 | 19 | 39 | 54 | 11 | 15 | 26 | 36 | 17 | 15 | 32 | 44 | 18 | 17 | 35 | 49 |
| 37 (4) | *Min. 6th*—Min. 3rd | 27 | 20 | 47 | 65 | 16 | 24 | 40 | 55 | 28 | 27 | 55 | 76 | 23 | 25 | 48 | 67 |
| 38 (3) | *Octave*—Prf. 5th | 19 | 19 | 38 | 53 | 17 | 21 | 38 | 53 | 19 | 27 | 46 | 64 | 25 | 26 | 51 | 71 |
| 39 (2) | Maj. 7th—*Aug. 4th* | 9 | 14 | 23 | 32 | 19 | 17 | 36 | 50 | 9 | 9 | 18 | 25 | 9 | 10 | 19 | 26 |
| 40 (1) | Maj. 7th—*Maj. 3rd* | 12 | 16 | 28 | 39 | 17 | 18 | 35 | 49 | 6 | 8 | 14 | 19 | 14 | 13 | 27 | 37 |
| 41 (25) | *Maj. 6th*—Min. 6th | 24 | 25 | 49 | 68 | 19 | 20 | 39 | 54 | 25 | 31 | 56 | 78 | 27 | 24 | 51 | 71 |
| 42 (24) | Maj. 2nd—*Octave* | 6 | 7 | 13 | 18 | 2 | 2 | 4 | 6 | 5 | 5 | 10 | 14 | 5 | 10 | 15 | 21 |
| 43 (23) | *Maj. 3rd*—Prf. 4th | 11 | 16 | 27 | 37 | 13 | 15 | 28 | 39 | 19 | 10 | 29 | 40 | 18 | 13 | 31 | 43 |
| 44 (22) | Min. 7th—*Maj. 6th* | 11 | 10 | 21 | 29 | 11 | 11 | 22 | 30 | 4 | 10 | 14 | 19 | 7 | 7 | 14 | 19 |
| 45 (21) | *Min. 6th*—Dim. 5th | 17 | 19 | 36 | 50 | 17 | 21 | 38 | 53 | 18 | 20 | 38 | 53 | 10 | 16 | 26 | 36 |
| 46 (20) | Dim. 5th—*Min. 3rd* | 20 | 11 | 31 | 43 | 11 | 22 | 33 | 46 | 10 | 9 | 19 | 26 | 18 | 20 | 38 | 53 |
| 47 (19) | Min. 2nd—*Maj. 7th* | 20 | 21 | 41 | 57 | 16 | 19 | 35 | 49 | 21 | 21 | 42 | 58 | 24 | 25 | 49 | 68 |
| 48 (18) | *Prf. 4th*—Min. 7th | 4 | 5 | 9 | 12 | 8 | 5 | 13 | 18 | 7 | 8 | 15 | 21 | 8 | 7 | 15 | 21 |
| 49 (17) | *Min. 6th*—Maj. 2nd | 8 | 12 | 20 | 28 | 17 | 9 | 26 | 36 | 8 | 12 | 20 | 28 | 16 | 10 | 26 | 36 |
| 50 (16) | *Maj. 3rd*—Dim. 5th | 17 | 17 | 34 | 47 | 20 | 21 | 41 | 57 | 12 | 15 | 27 | 37 | 18 | 17 | 35 | 49 |

the "consonance" test. The figures shown in the remaining columns are based upon results of the "preference" tests and have been calculated on the same basis as those just cited.

As previously indicated, reliability coefficients based on gross scores secured for the Seashore test are so low that the judgments appear to be largely the results of chance. Thus it would seem that consonance is too variable a phenomenon to be dealt with scientifically—that as far as science is concerned it is not a "phenomenon" at all. However, this extreme view is invalidated by the results shown in table 12, where the reliability coefficients are based on error-frequency per combination for the 36 subjects. The coefficient .87, for example, is secured by correlating the values in columns 3 and 4 in table 11; the other coefficients shown in table 12 are derived in like manner from values shown in table 11. The average reliability coefficient based upon these error-frequencies per combination for the several consonance series is .80, while the analogous correlation for the preference judgments is .90. These average coefficients indicate a very high *group consistency* in regard to the relative difficulty of the paired-interval comparisons. This discrepancy in "reliability" is doubtless due to the difference in the number of items involved in the two instances. For example, the reliability coefficient of .53 shown in table 9 is based on 36 cases—the scores of 36 subjects, whereas the reliability coefficient of .87 shown in table 12 for the same series is based on 50 items, each of which represents an *average* for 36 subjects. Thus it is obvious that consonance is not the absolutely irregular and fortuitous phenomenon that the usual type of "reliability coefficients," based

TABLE 12

*Showing reliability coefficients based on the error-frequencies (shown in table 11) per combination for 36 subjects.*

| | Consonance | | Preference | |
|---|---|---|---|---|
| Correlated | *R* | P.E. | *R* | P.E. |
| 1–2 | .87 | .03 | .93 | .01 |
| 1–3 | .73 | .05 | .91 | .02 |
| 1–4 | .87 | .03 | .89 | .02 |
| 2–3 | .73 | .05 | .89 | .02 |
| 2–4 | .82 | .04 | .90 | .02 |
| 3–4 | .79 | .04 | .91 | .02 |
| Average | .80 | | .90 | |

on gross scores, would seem to indicate. A "score" on a "consonance test" is a value whose precise significance no one knows. As previously indicated, it is subject to so many chance factors that it cannot be regarded as a reliable index of a subject's capacity for making this type of discrimination. However, when the reliability coefficients are based on a large number of actual responses to tonal stimuli rather than upon "scores" of individual subjects they indicate a high degree of general consistency for the group as a whole.

The failure to distinguish between these two ways of regarding "reliability" seems to be partly responsible for the disagreement which exists between the views of Heinlein and Larson with respect to the reliability of the Seashore test. In 1925 Heinlein reported a study (8) in which he concluded that a "score" made on the Seashore Consonance Test cannot be regarded as a true index of a subject's ability to judge relative consonance. This conclusion was based chiefly upon the fact that reliability coefficients secured for this test are much lower than those generally required for prediction in individual cases. However, in 1929 Larson published a monograph (14) which contained the results secured for six groups of subjects, both musically trained and untrained, who were given the Seashore test. In this study, Larson concluded that the paired-interval comparison method is adequate for testing consonance. Her conclusion was based chiefly upon an analysis of the results secured for each of the fifty paired-interval comparisons contained in the Seashore test. According to Larson "a high correlation" between the total number of errors per pair was obtained for the three applications of this test.[1] Thus, it seems evident that Larson thinks of "reliability" in terms of paired-interval consistency,[2] whereas Heinlein has in mind a "reliability" based on correlations of gross scores. Hence, for reasons already stated it is only natural that these two investigators should

[1] No specific coefficients are given by her.

[2] This is manifestly the case since the correlation coefficients secured by Larson (.63 ± .022 and .65 ± .024) do not differ greatly from those obtained by other investigators, and these have generally been held to be too low for satisfactory reliability.

come to different conclusions concerning the "reliability" of the Seashore test. Larson is apparently correct in holding that consonance discrimination is not an absolutely irregularly conditioned process, but she appears to be guilty of equivocation with respect to the meaning of "reliability." She does not appear to be warranted in concluding that a single "score" is a reliable index of consonance discriminability. When the judgments of a large group of subjects are "pooled," in a sense, the chance errors of the individuals apparently tend to cancel each other, and the "reliability" of such composite judgments is consequently high. But when the index of consistency is based upon the correlation of individual raw "scores," such errors affect the correlation between successive "tests" directly, with the result that these "reliability" indices are too low to be of much significance.

### 2. *Analytical study of the influence of difficult pairs of intervals on reliability*

It is possible that both the accuracy and the consistency of consonance discrimination vary with the difficulty of the comparisons required. That is, if the discriminations called for are such that the correct judgment is fairly obvious in each instance, the judgments might be expected to vary little from one presentation to another. However, as the difficulty of a series is increased the judgments of the subjects are made with less certainty and confidence and hence might be expected to show less consistency than in the case of relatively easier comparisons.

In order to check the effects of "difficulty" of the test series upon the subjects' consistency the Seashore test was divided into four parts on the basis of the difficulty of the paired-intervals. The first part was composed of the ten easiest pairs of intervals, according to the error percentages shown in column 10 of table 11. This part included those pairs whose error percentage ranged from 6 to 22 per cent (inclusive), and was termed the *Very Easy* group. The second or *Fairly Easy* group included pairs whose error percentage ranged from 36 to 44 per cent. The third or *Difficult* group was composed of pairs having an

TABLE 13

*Showing the means, in terms of per cent correct, of the four 10-unit sub-tests.*

| | Consonance Tests | | | | Preference Tests | | | |
|---|---|---|---|---|---|---|---|---|
| Series | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Sub-test 1 | 86 | 84 | 85 | 86 | 90 | 89 | 86 | 81 |
| Sub-test 2 | 66 | 64 | 60 | 65 | 63 | 64 | 62 | 66 |
| Sub-test 3 | 56 | 54 | 52 | 49 | 62 | 56 | 54 | 52 |
| Sub-test 4 | 39 | 37 | 43 | 35 | 39 | 38 | 40 | 39 |

error percentage of 44 to 54 per cent. The fourth or *Very Difficult* group included pairs whose error percentage ranged from 55 to 76 per cent. After thus selecting the pairs of intervals to constitute the four sub-tests, a " score " for each of these was computed for each subject. The per cent of correct judgments for the ten pairs of intervals constituted the " score " of the subject on each of the sub-tests. The averages for each of these sub-tests are shown in table 13. The difference in the size of the means between each of these sub-tests is sufficient to indicate the reality of the distinction between the several series in difficulty. Sub-tests 2 and 3 of the " preference " series show a smaller difference than do the other comparisons, a fact which may be due partially to the subdivision of the combinations on the basis of " consonance " judgments.

Table 14 indicates in some measure the degree to which the reliability of consonance judgments is influenced by the difficulty of the comparisons required. This table shows the reliability coefficients and the total number of judgment reversals, with their corresponding per cents, for each sub-test. By a judgment

TABLE 14

*Showing the reliabilities of the four sub-divisions of the Seashore Consonance Test.*

| Sub-tests | Series Correlated | *R* | P.E. | Average *R* | Judgment Reversals No. | Judgment Reversals Pct. |
|---|---|---|---|---|---|---|
| 1. Very Easy | 1–2<br>3–4 | .31<br>.37 | .10<br>.10 | .340 | 140 | 19.4 |
| 2. Fairly Easy | 1–2<br>3–4 | .46<br>.15 | .09<br>.11 | .305 | 269 | 37.3 |
| 3. Difficult | 1–2<br>3–4 | .33<br>.24 | .10<br>.11 | .285 | 282 | 39.1 |
| 4. Very Difficult | 1–2<br>3–4 | .38<br>.23 | .10<br>.11 | .305 | 269 | 37.3 |

reversal is meant simply that one time the subject records that a given pair is more consonant than another pair, whereas at another time he judges it to be less consonant. Obviously, the maximum number of judgment reversals which can be made by 36 subjects on four applications of a ten-unit sub-test, such as those referred to in table 14, is 720.[1] Hence Group 1 of this table, with a total of 140 judgment reversals, has the corresponding per cent, 19.4.

Heretofore it has been customary to think of the reliability of consonance judgments in terms of reliability coefficients secured by correlating the scores (in terms of per cent correct) of the subjects made on different applications of a test. However, this method has been found to be very unsatisfactory, since often two applications of a test yield a relatively high number of judgment reversals, yet the reliability coefficient indicates greater consistency in judgment than appears for other less fluctuating series. For example, the first and second presentations of the *Very Easy* group shown in table 14 have an $R$ of .31 and 70 judgment reversals, while the first and second presentations of the *Very Difficult* group have an $R$ of .38, although the number of judgment reversals totals 125. Now one would naturally suppose that the reliability coefficient for consonance judgments would vary inversely with the number of times the subjects reverse their decisions, but we have here an instance in which the size of the reliability coefficient varies directly with the number of judgment reversals. This rather unexpected result is due to the fact that in one instance the index of reliability is based on the number of times the subjects reverse their decisions, irrespective of the correctness or incorrectness of the latter, whereas in the other instance it is based on the correspondence between the gross scores in terms of per cent correct. The following example will serve to show how these two methods of approach may lead to contradictory results.

Let us suppose that a consonance test composed of twenty comparisons is presented to a subject twice, the latter making a

[1] That is, when the number of judgment reversals are calculated for the first two and the last two applications.

score of 90 per cent correct upon each presentation. Now, from the point of view of " scores " this represents perfect consistency. However, if the two errors made upon the first presentation are different from those made on the second, we have four judgment reversals notwithstanding a perfect correspondence in scores. On the other hand, let us suppose that the subject made a score of 90 on the first presentation and 75 on the second. From the point of view of scores this represents a lower degree of consistency than was present in the above case. However, if two of the five errors made on the second presentation of the test duplicate those made upon its first presentation, we have only three judgment reversals. Thus it can be seen that two entirely different indices of reliability are available in these two types of values. Apparently, the number of judgment reversals made by a group on any such series of comparisons is a more accurate index of consistency than is a reliability coefficient based on " scores " in terms of per cent correct. The number of judgment reversals increases directly with each inconsistent response, whereas the " per cent correct " score permits compensation of a reversal scored as an error by another reversal scored as " correct." The foregoing discrepancy in reliability furnishes further evidence in support of the view previously implied, namely, that a consonance test " score " is a vague general index which is based upon diversely conditioned responses and which only imperfectly indicates the subject's capacity for consistent discrimination. In view of this fact the per cents of judgment reversals shown in table 14 rather than the *Rs* secured by correlating the gross " scores " of the subjects, are considered to be more significant.

Examination of table 14 shows that although the average *R*s secured for the four sub-tests are practically the same, the two subdivisions composed of the relatively easy pairs of intervals are more reliable, in terms of judgment reversals, than those composed of more difficult pairs. The average percentage of judgment reversals for sub-tests 1 and 2 is 28.3, as compared with 38.2 for sub-tests 3 and 4. The most striking difference in reliability occurs between sub-tests 1 and 3, the subjects reversing

their decisions almost twice as many times on the latter sub-test as on the former. These comparisons indicate that subjects are more consistent when the comparisons required are such that the "correct" decision is fairly obvious in each case than when the pairs are of such difficulty as to make the subject uncertain as to the correct judgment. However, it must not be thought that the most difficult pairs of intervals (those having the highest error-frequency) give rise to the greatest inconsistency. On the contrary, table 14 shows that the fourth sub-test which was composed of very difficult pairs had a slightly lower percentage of judgment reversals (37.3) than the third sub-test (39.1) which was composed of less difficult pairs. Although this difference in consistency is statistically insignificant, the slightly superior reliability of the *Very Difficult* over the *Difficult* sub-test was perhaps to have been expected. The former sub-test was composed of pairs having an error percentage of from 55 to 76 per cent, whereas the latter sub-test was composed of pairs with an error percentage of from 44 to 54 per cent. In the case of the *Very Difficult* group the pairs presented such difficulty that the subjects did not have an even chance of making the correct judgment. That is to say, certain factors seemed to operate in favor of the incorrect judgment, thus making for a constant error. This being the case, it is easy to understand why the *Very Difficult* group, which was probably influenced by such factors, tended to be slightly more reliable than the *Difficult* group which was affected to a less degree by such factors.

A comparison of the results in table 14 with those shown in table 9 reveals that, in general, the reliability coefficients of the ten-unit sub-tests are much lower than those of the fifty-unit test. The average reliability for the sub-tests (all repetitions included) is .31 as against .52 for the fifty-unit test. In view of the work of Lanier[1] it is likely that this disparity in reliability

[1] Lanier (13, p. 96) found that artificially constructed ten-unit tests, based upon Seashore Consonance Test results, gave an average $R$ of .158. There was a fairly regular increase in the reliability coefficient with increased length of the test, up to thirty units (paired-intervals), where the average $R$ was .510. The increase in the size of the reliability coefficients upon increasing the number of test items was considerably less than that predicted with the Spearman-Brown formula.

is due to the difference in the number of units of which the two types of tests are composed. In order to remedy, partially, this defect of too few test items, in the case of the sub-tests, the original fifty-unit test was scored on the basis of the twenty easiest and the twenty most difficult pairs of intervals. Two sub-tests, each of twenty paired-interval comparisons, were thus constituted, and the per cent of correct judgments for each part was computed for all subjects. The average mean (based on the four applications) for the group composed of the twenty easiest pairs was 80 per cent correct for the consonance judgments and 85 per cent correct for the preference judgments. The average mean for the group composed of the twenty most difficult pairs was 46 per cent correct for the consonance judgments and 48 per cent correct for the preference judgments. Thus, there is sufficient disparity between the two sets of averages to insure the existence of two distinct groups in point of difficulty.

Table 15 shows that the effect of lengthening the sub-tests was to raise the reliability coefficients considerably, although they still fall short of satisfactory reliability. The sub-test composed of the twenty easiest pairs of intervals has an average *R* of .595, while the sub-test composed of the twenty most difficult pairs has an average *R* of .35. The *R* for the twenty easiest pairs is higher than the *R* of .52 secured for the entire fifty-unit test. The analogous values for the " preference " series were .575 and .240. Table 15 shows, in general, that for both the consonance and preference series, the sub-test composed of the easier pairs of intervals has much higher reliability coefficients than the sub-test composed of the more difficult pairs. These results confirm those

TABLE 15

*Showing the reliabilities of sub-tests composed of the 20 easiest and the 20 most difficult pairs of intervals in the Seashore Consonance Test.*

| | | Consonance | | Preference | |
|---|---|---|---|---|---|
| Sub-test | Series Correlated | *R* | P.E. | *R* | P.E. |
| 20 Easiest | 1–2 | .54 | .08 | .60 | .07 |
| | 3–4 | .65 | .06 | .55 | .08 |
| Average | | .595 | | .575 | |
| 20 Most Difficult | 1–2 | .42 | .09 | .32 | .10 |
| | 3–4 | .28 | .10 | .16 | .11 |
| Average | | .350 | | .240 | |

of the previous analysis, which was based on per cents of judgment reversals, in that they indicate that subjects are more consistent when making easy comparisons than when making relatively difficult ones. By way of explanation of this fact, it seems probable that the element of chance is more prominent in the latter case than in the former, and more than offsets any constant error which in itself would tend to make for consistency.

### 3. *The reliability of consonance judgments in separate tests of varying difficulty*

The above results were based upon the manipulation of data secured with the Seashore Consonance Test, in which easy and difficult combinations were indiscriminately mixed. It seemed desirable to construct several separate series of combinations varying in difficulty in order to isolate if possible the effects of difficult pairs of intervals upon relatively easy combinations. This particular matter, *i.e.*, the influence of difficult pairs of intervals within a consonance series upon the judgment of less difficult pairs, constitutes an important problem which has, for the most part, been ignored by the writers on consonance perception. It has apparently been assumed by Seashore and others that consonance judgments represent responses confined entirely to certain specific pairs of intervals. That is to say, the subject is supposed to react to any given paired-interval within a series irrespective of the tonal stimuli which have preceded it. However, this appears to be an unwarranted assumption; it is a matter of common knowledge that the presence of very difficult items within a test often proves discouraging to the subject, and thus lessens his efficiency with respect to other less difficult items. In other words, one's mental attitude is an important factor in any work requiring discrimination. Now, there seems to be no valid reason why consonance discrimination should constitute an exception to this general rule. On the contrary, overlapping attitudes and sets engendered by difficult pairs of intervals undoubtedly have their effects upon the judgment of other less difficult pairs within the series. In view of this fact we should find variations in the accuracy and consistency of consonance

judgments with the introduction of difficult combinations into the series of paired-interval comparisons.

As previously stated, in order to investigate the above problem four tests (shown in tables 1, 2, 3, and 4) composed of twenty, twenty-five, thirty, and twenty units respectively were constructed. Test 1 was composed of twenty easy pairs of intervals. Test 2 was composed of twenty-five units, the twenty easy pairs of Test 1 and five difficult pairs interspersed at regular intervals, the latter pairs being used merely as distractors. Test 3 was composed of thirty units, the same twenty easy pairs and ten difficult pairs introduced at regular intervals. Tests, 1, 2, and 3 were scored on the basis of the same twenty easy pairs of intervals contained in each, the " difficult " comparisons being omitted in computing the several " scores." Test 4 was composed of twenty difficult pairs of intervals, and was scored on the basis of the per cent of these judged correctly. Table 16 gives the mean scores in terms of per cent correct, the standard deviations and the reliability of the differences between certain of the means, for the two applications of the above four tests. For example, the mean for the first presentation (noted as 1a) of Test 1 is 79.24 per cent, while the mean for the first presentation (noted as 2a) of Test 2 is 74.35 per cent.

Examination of table 16 shows that the addition of five difficult pairs to the twenty easy combinations resulted in more of the easy pairs being missed than was the case when the series consisted entirely of easy pairs. The average for the first presentation of Test 2 falls 4.89 points, in terms of per cent correct,

TABLE 16

*Showing means in terms of per cent correct, standard deviations, and the reliability of the differences for the two presentations of the four series, each varying in difficulty, for 39 subjects.*

| Series | 1a | 1b | 2a | 2b | 3a | 3b | 4a | 4b |
|---|---|---|---|---|---|---|---|---|
| Mean | 79.24 | 83.46 | 74.35 | 78.20 | 82.17 | 78.08 | 54.75 | 61.00 |
| S.D. | 14.00 | 10.30 | 16.40 | 11.45 | 12.24 | 12.05 | 9.75 | 11.60 |
| Means Compared | 1a–2a | 1b–2b | 1a–3a | 1b–3b | 2a–3a | 2b–3b | | |
| Diff. / Sigma Diff. | 1.42 | 2.15 | .99 | 2.13 | 2.40 | .04 | | |

Note: The letters *a* and *b* refer to the first and second presentations of a series, respectively.

below the average of the first presentation of Test 1, while the average for the second presentation of Test 2 is 5.26 points less than that for the second presentation of Test 1. The Diff./Sigma Diff. values for these differences are 1.42 and 2.15, respectively. Thus it would seem that the initial effect of the introduction of difficult pairs of intervals into a standard series is to reduce the average score on the latter. The mean for the first presentation of Test 3 is higher than that for either presentation of Test 2. Again, when both presentations of Tests 1 and 3 are taken into account the two tests appear to present about the same degree of difficulty, the former test having an average of 81.35 per cent correct as compared with 80.12 per cent for the latter. By way of explanation of these facts, it is possible that by the time Test 3 was given some of the subjects had become adapted to the distractors so that the latter did not reduce the subjects' efficiency. The means for the two applications of Test 4, which was composed entirely of difficult pairs, indicate that many of the so-called consonance judgments were little more than mere guesses. In this connection it should be recalled that the twenty pairs constituting Test 4 represented 40 per cent of the paired-intervals of the Seashore test. This may account in some measure for the low reliability which has generally been obtained for the latter test. The fact that it is possible to construct, from pairs contained in the Seashore test, a series of such difficulty that the results secured for it will be little more than mere guesses, suggests that the low reliability of the former test is due partly to these difficult test items. However, in view of the error-frequencies shown in table 11, the possibility of constructing such a test as the above should not be surprising. Column 10 of table 11 shows that 22 of the 50 paired-intervals of the Seashore test have an error percentage ranging from 40 to 76 per cent inclusive, giving an average of 54.22 per cent. Any test composed of so many difficult test items could scarcely be expected to yield very high reliability.

Additional data concerning the relation between the reliability of a series and the difficulty of the pairs of intervals of which it is composed are presented in table 17. This table also shows the

TABLE 17

*Showing the reliabilities for the four tests of varying difficulty in which 39 subjects were used.*

| Series Correlated | *R* | P.E. | Judgment Reversals No. | Pct. |
|---|---|---|---|---|
| 1a–1b | .15 | .11 | 209 | 26.7 |
| 2a–2b | .65 | .06 | 230 | 29.4 |
| 3a–3b | .39 | .09 | 209 | 26.7 |
| 4a–4b | .46 | .08 | 302 | 38.7 |

Note: The letters *a* and *b* refer to the first and second presentations of a series, respectively.

effect of the presence of difficult pairs of intervals within a series upon the consistency of the judgments with respect to less difficult pairs. It is evident from an examination of the above table that the number of judgment reversals and the reliability coefficients per test do not agree. According to its reliability coefficient, Test 1 which has an *R* of .15 is the least reliable of the four tests, whereas according to the number of judgment reversals it shares with Test 3 the highest reliability, having only 209 judgment reversals. It is possible that this low *R* is due to the ease of the test, since a series which is too easy makes for a rather homogeneous group and low "reliability." However, we have already explained (*supra,* pp. 58–59) how this inconsistency may arise, and have indicated that the number of judgment reversals is probably a more accurate index of reliability than a reliability coefficient based on the scores in terms of per cent correct.

The data secured for the foregoing tests indicate further the correspondence between the difficulty of the comparisons and the consistency of the subjects' judgments. Table 17 shows that Test 1, which has an average mean of 81.35 per cent correct for its two presentations has a total of 209 (or 26.7 per cent) judgment reversals, while Test 2, which contained five difficult pairs used as distractors has an average mean of 76.27 per cent and 230 judgment reversals. This indicates that the introduction of the distractors into the series not only lowered the efficiency of the subjects with respect to the easy pairs, but also caused them to be less consistent in their judgments. According to the average

mean, Test 3 (although containing ten distractors) was of approximately the same difficulty as Test 1, having an average mean of 80.12 per cent. As previously mentioned, it is possible that by the time this test was given the subjects had become adapted to the distractors so that the latter did not greatly affect the subjects' judgments in the case of the easy pairs. The important point to notice here, however, is that Tests 1 and 3 which offer practically the same degree of difficulty, as defined by performance, show the same reliabilities, each having 209 judgment reversals. Test 4, which was composed entirely of difficult pairs, has the lowest reliability, the subjects having made 302 judgment reversals. This result, however, in no way contradicts the statement previously made that a very difficult consonance test may be more reliable than one of slightly less difficulty. In the case of the present study, Test 4 was the only test which could really be regarded as difficult.

In conclusion, the results secured for the four tests varying in difficulty agree with those obtained for the four sub-tests which were artificially constituted from the Seashore test results (see table 14). On the basis of these two studies the following conclusions may be drawn:

1. The reliability of a consonance series tends to vary somewhat inversely with the difficulty of the comparisons called for.

2. Difficult pairs, particularly those of moderate difficulty, make for unreliability in that they bring about an attitude of guessing on the part of the subjects.

3. The initial effect of the presence of difficult pairs of intervals within a consonance series is to reduce the efficiency and consistency of the subjects with respect to less difficult pairs.

4. Certain very "difficult" pairs of intervals show two opposite trends: first, the judgments seem to be conditioned by an attitude of guessing on the part of the subjects which makes for unreliability, and second, certain constant errors occur consistently and so make for a lower percentage of judgment reversals. These two trends were mentioned at the beginning of the present chapter. Further experimentation must be made

before the precise effects of these two types of " difficulty " upon " reliability " can be determined.

5. The low reliability of the Seashore Consonance Test is probably due in part to its inclusion of a high percentage of difficult test items. Approximately 50 per cent of the paired-intervals contained in this test present such difficulty that the judgments secured for them seem to be largely the result of chance factors. Furthermore, the presence of these difficult pairs of intervals within the test probably lowers the accuracy and consistency of the subjects with respect to the other less difficult test items. In view of the composition of this test it is hardly to be expected that even the nicety of instructions such as Larson has insisted upon could secure for it a satisfactory reliability.

# CHAPTER VI

## The Influence of Affective-Tone on Consonance Judgments

The claim has been made that due to affective-tone there is hardly such a thing as a comparison of two sets of intervals purely on the basis of consonance (20, p. 31). "Affective-tone" is a general term, used here to denote the pleasantness or unpleasantness which may characterize an interval. It may be due to progression, resolution, or merely to the intrinsic quality of an interval. By "progression" is meant a motion from note to note or from chord to chord. The term is sometimes used to define the general aspect of a more or less extended group of modulations with reference to the order of their succession. Thus, when a subject hears a given chord followed by another, and the two are so related to each other that the result seems to be the beginning of a melody he is likely to be influenced by this melodic relationship. "Resolution," according to Grove (5), consists in the process of relieving dissonances by succeeding consonances. However, we must regard the terms consonances and dissonances relatively. For example, in passing from the Perfect Fourth (c′f′) to the Major Third (c′e′) one experiences a feeling of repose or satisfaction. The last chord seems to answer the question which was raised by the first, and we desire no further chord for finality. But it is generally agreed that the Major Third is less consonant than the Perfect Fourth, and so in this instance resolution does not consist in the process of relieving dissonance by a succeeding consonance. We are not, however, so much interested in calling attention to the fact that Grove's definition of resolution has certain exceptions, as in indicating the general psychological nature of resolution, which is a feeling of satisfaction or repose accompanying certain intervals when they have been preceded by other intervals. By the "intrinsic quality" of an interval we have reference to its rough-

ness or smoothness. Certain intervals give rise to beats which impart a rough jarring quality to the tonal complex, making it unpleasant, whereas other intervals which are relatively free from beats possess a smoothness which is agreeable.

While the foregoing terms have specific connotations which come chiefly from the field of music, their psychological effects in relation to the perception of consonance are similar. That is, they make for either a pleasant or an unpleasant tonal experience. The factor of progression may have either one of two effects. If the second interval is related to the first in such a manner as to remind the listener of some familiar melody, it will probably be heard as pleasant. On the other hand, if the first chord causes the listener to expect a progression which the second chord fails to provide the latter is felt as a disappointment and as such has an unpleasant effect. In the case of resolution the effect is almost invariably pleasant, that is, the final chord seems to satisfy the listener. Now, as already indicated, it has often been alleged that the presence of such an element so colors the tone quality that one finds it almost impossible to disregard it, and make a "cognitive" comparison purely on the basis of relative consonance. Affective-tone due to the intrinsic quality of an interval may be either pleasant or unpleasant, according as the interval is smooth or rough. If an interval such as the Minor Second is compared with the Major Sixth, the listener will probably be biased in favor of the correct judgment since the latter interval is not only the more consonant but also the more pleasing of the two.

Almost all investigators of the problem of consonance have held that affective-tone influences "consonance" judgments. In one of the earliest experimental studies of consonance Moore (**18**, p. 50) made affective-tone the criterion of consonance. A later study of the same general problem by Malmberg (**17**, p. 126) is admittedly open to the criticism that the results were not free from the influence of affective-tone. Continuing the same type of investigation, Gaw (**4**, p. 140) met with the same difficulty, particularly in the case of undue preference for the Major Third. Guthrie and Morrill (**7**, p. 625) in ranking intervals on the basis

of both consonance and pleasantness found a high agreement between the two curves. In a somewhat recent investigation Guernsey (6) secured results which led her to regard the criterion *affective-tone* as superior to either *smoothness* or *fusion,* and she concluded that pleasantness and unpleasantness are the most legitimate criteria of consonance. On the other hand, Heinlein (8) claims that the presence of the factor of affective-tone in connection with the Seashore Consonance Test renders the latter invalid as a measure of consonance perception. Peterson (20, p. 31) seems to share this view when he says:

> "Tonic effects, preparations for resolutions, associations with this or that progression in familiar musical selections, etc., are almost inevitable and seem to be troublesome to many subjects. There is hardly such a thing as a purely independent comparison of two sets of intervals on their own degrees of consonance."

Larson (14, p. 62), however, partly on the basis of introspectional evidence given by her subjects, holds that harmonic progression is a negligible factor. There exist, then, at least three different views with respect to affective-tone; first, that it operates as a legitimate factor in determining consonance judgments; second, that it is a non-essential and disturbing factor; and third, that it is negligible.

In order to secure additional evidence on this question an analysis has been made of the relevant results obtained from the four "consonance" and the four "preference" tests. Certain data shown in tables 7, 9, and 12 were used. We have attempted to make a general study of the relation between consonance judgments and preference judgments by comparing the averages and reliabilities of the two series of tests, and also by correlating the two types of scores. In addition to these general comparisons a detailed analysis of the effects of order of presentation of intervals upon consonance discrimination has been made. This was accomplished by comparing the results (for both orders of presentation shown in table 11) on certain pairs of intervals when the subjects were instructed to disregard affective-tone and make a detached judgment on the basis of relative consonance with those secured for the same pairs when affective-tone was made the basis of the decisions.

1. *General comparisons of consonance and preference judgments*

The results of the general comparisons of "consonance" and "preference" judgments are shown in tables 7, 9, and 18. The mean scores in table 7 indicate that the consonance and the preference series are of about equal difficulty. The averages for the preference judgments are somewhat higher than those for the consonance judgments, although the differences are statistically insignificant. The correlations between the consonance and the preference judgments shown in table 18 indicate that there is approximately the same degree of correspondence between the respective tests of the consonance and of the preference series as exists between the several consonance tests. In the first case the average $r$ is .56, while the intercorrelation of the consonance test yields an average $r$ of .52.

The comparative reliabilities of the consonance and the preference judgments are shown above in table 9. The average reliability of the consonance judgments, when based on the scores, is .52, as compared with .60 for the preference judgments. When the reliability coefficients are based on the error-frequencies per combination shown in table 11 the superiority of the preference judgments is more marked, the latter having an average reliability of .90 as against .80 for the consonance judgments. These results seem to indicate that subjects are more consistent when recording their preferences than when attempting to make cognitive judgments of relative consonance. It is likely that in the case of the preference judgments a simpler, more constant basis for judging exists. The most natural response to the hearing of an interval is perhaps one of approval or disapproval.

TABLE 18

*Showing the coefficients of correlation between the four consonance and the four preference tests.*

| Series Correlated | $r$ | P.E. |
|---|---|---|
| 1–1 | .70 | .06 |
| 2–2 | .56 | .08 |
| 3–3 | .35 | .10 |
| 4–4 | .63 | .07 |
| Average | .56 | |

It is appreciably easier to maintain this attitude as a basis for judging through an entire series and from one series to another than to achieve consistency in the comparison of intervals on the basis of relative consonance. In the case of so-called consonance discrimination it is possible that the subject finds it very difficult, if not impossible, to disregard the affective aspect of certain combinations. As a result of this he probably does not use a single criterion or set of criteria from one series to another or within the same series, but judges some of the pairs on a cognitive basis, and others largely on an affective basis. In still other instances the decisions are probably determined by various chance factors. If such a fluctuation between a cognitive and an affective basis for judging actually obtains, it would doubtless account for much of the inaccuracy and inconsistency of so-called consonance discrimination.

### 2. *An analysis of the effect of order of presentation of intervals upon "consonance" and "preference" judgments*

The fact that the average mean for the consonance judgments was found to approximate to that for the preference judgments, and the further fact that consonance test scores and preference test scores correlate as highly as do consonance scores suggest the possibility that in some instances these supposedly two types of "discriminability" actually involve the same kind of psychological "set." That is to say, it is quite likely that many of the so-called consonance judgments were actually determined by the same factors (resolution, progression, etc.) which determined the preference judgments. Furthermore, if such factors as resolution, progression, etc., do have an effect upon consonance judgments, then certain types of judgment reversals can be expected when the order of presentation of the intervals is reversed. Thus a comparison has been made of the results (for both orders of presentation shown in table 11) on certain pairs of intervals when the subjects were instructed to make cognitive judgments on the basis of relative consonance with those secured for the same pairs when affective-tone was made the basis of the decisions. The data secured for these selected pairs of intervals

are shown in table 19. Table 11, upon which table 19 is based, shows the percentage of *errors* for each pair, whereas table 19 shows the percentage of *correct* judgments.

An analysis of the results secured for pairs 23 and 43 of the Seashore test affords us an opportunity to study two of the instances in which it has been alleged that affective-tone, due to resolution, influences the comparison of the intervals on the basis of consonance. In the case of pair 23, a subdominant to tonic relationship exists. In passing from the fourth to the third, in the same tonality, the psychological feeling is one of completeness or finality which is supposed to be relatively satisfying. Pair 43, which consists of the same intervals in reverse order, constitutes the passage from a dominant to a tonic position, and carries with it also a feeling of finality and satisfaction which is supposed to bias the subject in favor of the fourth. Thus, both pairs 23 and 43 constitute instances of "resolution." If consonance discrimination is influenced by this kind of harmonic relationship, then judgment reversals should be found for these pairs. That is, we should find the Major Third generally regarded as the more consonant interval in pair 23, and the Perfect Fourth the more consonant interval in pair 43. At first sight it would appear that the order of presentation had no effect

TABLE 19

*Per cent of correct judgments on selected pairs of intervals, showing the extent to which affective-tone influenced the judgments.*

| | | Consonance | | | Preference | | |
|---|---|---|---|---|---|---|---|
| Pair | Intervals | Series 1 and 2 | Series 3 and 4 | Av. | Series 1 and 2 | Series 3 and 4 | Av. |
| 23 | Prf. 4th—Maj. 3rd | 60 | 64 | 62.0 | 70 | 79 | 74.5 |
| 43 | Maj. 3rd—Prf. 4th | 63 | 61 | 62.0 | 60 | 57 | 58.5 |
| | Average | | | 62.0 | | | 66.5 |
| 3 | Prf. 5th—Octave | 83 | 89 | 86.0 | 82 | 68 | 75.0 |
| 38 | Octave—Prf. 5th | 47 | 47 | 47.0 | 36 | 29 | 32.5 |
| | Average | | | 66.5 | | | 53.7 |
| 7 | Maj. 3rd—Maj. 6th | 58 | 49 | 53.5 | 70 | 64 | 67.0 |
| 34 | Maj. 6th—Maj. 3rd | 43 | 43 | 43.0 | 35 | 35 | 35.0 |
| | Average | | | 48.2 | | | 51.0 |
| 8 | Maj. 3rd—Min. 6th | 60 | 61 | 60.5 | 33 | 36 | 34.5 |
| 33 | Min. 6th—Maj. 3rd | 70 | 79 | 74.5 | 86 | 83 | 84.5 |
| | Average | | | 67.5 | | | 59.5 |
| 28 | Octave-Maj. 3rd | 40 | 60 | 50.0 | 26 | 18 | 22.0 |

upon the "consonance" judgments since table 19 shows an average of 62 per cent "correct"[1] for each of the above pairs. However, an analysis of the data obtained for pairs 23 and 43, for both the "consonance" and the "preference" series, does not bear out this initial impression. The results secured for pair 23 show that the subjects "preferred" the Major Third to the Perfect Fourth in 74.5 per cent of the cases, and since, according to the norm adopted for scoring the third is the more consonant interval, the subjects have an average of 74.5 per cent "correct" for this pair. In the case of pair 43 the Perfect Fourth, which completed the resolution, was preferred in 41.5 per cent of the cases, whereas it was preferred in only 25.5 per cent of the cases for pair 23. Since, according to our norm, it is less consonant than the Major Third the subjects averaged 58.5 per cent "correct" in the case of pair 43. The foregoing results indicate that although the order of presentation influenced the preferences it did not do so sufficiently to cause the Perfect Fourth to be preferred in the majority of cases for pair 43. It is possible that pair 43 is subject to two counter influences. Resolution probably tends to bias the subject in favor of the Perfect Fourth whereas the pleasingness of the Major Third (which general experimental results [4, p. 140] have indicated as being more pleasing than the Perfect Fourth) tends to bias the subject in its favor. Apparently, the intrinsic pleasingness of the third more than offsets the effects of resolution due to the third being followed by the fourth. This suggests that preference for the third was the determining factor in the majority of cases for both the consonance and the preference judgments for both of the above pairs. It would seem to be more than a mere coincidence that the Major Third was regarded as the more consonant and also as the more pleasing of the two intervals by the majority of subjects. The interpretation of this correspondence as indicative

[1] The Seashore norms have been adopted merely as a matter of convenience, *i.e.*, for the purpose of providing a standard by which the consistency of the judgments might be most conveniently checked. Thus, our use of such terms as "correct," "incorrect," "error," "efficiency," etc., carries with it only a reference to these arbitrarily adopted norms which may or may not be correct in certain instances.

of the influence of affective-tone upon consonance judgments is further supported by the fact that (notwithstanding Seashore's norm) the Major Third is regarded by most authorities as more pleasing but less consonant than the Perfect Fourth.

An analysis of the data secured for pairs 3 and 38 furnishes further evidence with respect to the probable influence of order of presentation of intervals constituting a pair upon the consonance judgments secured. If such judgments are influenced by the order of presentation, then we should find widely divergent results for pairs 3 and 38 of the consonance series. Examination of table 19 shows such to be the case. The Octave is regarded as more consonant than the Perfect Fifth in 86 per cent of the cases when preceded by the latter interval, whereas when followed by the Perfect Fifth it is regarded as the more consonant in only 47 per cent of the cases. A corresponding variability is shown for the preference series. When the Octave follows the Perfect Fifth it is regarded as the more pleasing interval in 75 per cent of the cases, whereas when the order of presentation is reversed the Octave is regarded as the more pleasing interval in only 32.5 per cent of the cases (which means that the Perfect Fifth was regarded as the more pleasing in 67.5 per cent of the cases).[1] The most important point to keep in mind is the manner in which the consonance and preference series vary together. In general, when the Octave is regarded as more pleasing than the Perfect Fifth it is judged to be the more consonant, whereas when it is regarded as less pleasing than the Perfect Fifth it is judged to be the less consonant of the two. The foregoing results show that both consonance and preference judgments are influenced to a considerable degree by the order of presentation of the intervals. Furthermore, the fact that the frequency with which the Octave is judged more consonant than the Perfect Fifth tends to correspond to the frequency with which it is regarded as the more pleasing interval suggests that these supposedly two types of reaction often involve the same "set." Thus it is quite likely

[1] This preference for the Perfect Fifth when preceded by the Octave can perhaps be accounted for by the fact that we have here a passage from the dominant to the tonic through the melodic skip in the lower tones which causes the latter interval to be heard as the more satisfying combination.

that, in many cases,[1] the so-called consonance judgments are merely individual preferences based on the relative pleasingness of the intervals.

Pairs 7 and 34 are composed of the Major Third and the Major Sixth, each of which is generally regarded as a pleasing interval. Because of this similarity in affective quality the results secured for these pairs should indicate rather decisively whether or not the order of presentation affects either consonance or preference discrimination. According to the norm adopted, the Major Sixth is the more consonant of the two intervals. Table 19 shows that the sixth, when preceded by the third, is preferred in 67 per cent of the cases, and judged as more consonant than the third in 53 per cent of the cases. When the order of presentation is reversed, however, the sixth is preferred to the third in only 35 per cent of the cases. Corresponding to this shift in preferences, with the change in order of presentation, there is a reversal of judgment with respect to the relative consonance of the two pairs, the sixth being regarded as the more consonant interval in only 43 per cent of the cases. Although this shift in the consonance judgments coincident with the change in order of presentation is not as great as that for the preference judgments, the fact that fewer "errors" were made when the more consonant interval was regarded as the more pleasing interval suggests that some of the subjects were conditioned by the same factor in both types of judgments.

The effects of the order of presentation upon consonance discrimination are clearly shown by the results secured for pairs 8 and 33. Pair 8 consists of the Major Third followed by the Minor Sixth; in pair 33 the same intervals are presented in reverse order. In the case of the former pair, the Major Third was preferred in only 34.5 per cent of the cases, although it was regarded as the more consonant in 60.5 per cent of the cases. These facts seem to indicate that some of the subjects were able to disregard their preferences and make relatively detached judg-

[1] The fact that the average for the consonance judgments for pair 38 is 14.5 points higher than that for the preference judgments for the same pair indicates that some of the subjects made cognitive judgments.

ments on the basis of consonance, since otherwise the per cent "correct" would have been much lower. However, when the order of presentation of the intervals is reversed the Major Third is preferred to the Minor Sixth in 84.5 per cent of the cases, and judged to be the more consonant interval in 74.5 per cent of the cases. Thus, when by reason of a change in its order of presentation the third becomes the more pleasing interval it also comes to be regarded as the more consonant one in approximately 20 per cent more of the instances.

In the case of pair 28, we have a comparison of the most consonant of all the intervals, the Octave, with the Major Third which is generally regarded as one of the two[1] most pleasing intervals. The results secured for this pair show that the Octave was preferred to the Major Third in only 22 per cent of the instances, and regarded as the more consonant interval in 50 per cent of the cases. In the light of our foregoing analyses, these facts seem to indicate: first, that a few of the subjects were able to disregard their feelings and make cognitive judgments, since otherwise the per cent "correct" for the consonance judgments would have approximated more nearly to that for the preference judgments; and second, that in the majority of cases the affective quality of the Major Third constituted such a bias in favor of the interval that, *statistically,* the outcome for the group was no better than that which would have been the result of mere chance.

Our general comparisons of consonance and preference judgments showed a high degree of correspondence between these supposedly two types of discriminability. That is to say, although the preference judgments tended to be somewhat more "accurate" and more consistent (in terms of scores and reliability coefficients) than so-called consonance judgments, the differences in many instances were so slight as to suggest that the "sets" were about the same in both cases. This conclusion has apparently been verified by the analysis just made of the effect of order of presentation of intervals upon consonance and preference

[1] The Major Third and the Major Sixth appear to be the most pleasing of all the intervals of the diatonic scale.

judgments. In the latter study it was shown that both the so-called consonance judgments and preference judgments vary with the order of presentation of the intervals, and more specifically that the frequency with which an interval is regarded as the more consonant one of a pair tends to vary directly with the frequency with which it is regarded as the more pleasing interval. Notwithstanding that consonance discrimination is supposed to be a cognitive rather than an affective process, these facts indicate rather strongly that it is often influenced by such affective factors as harmonic progression, resolution, etc.[1] This means that many of the so-called consonance judgments are not really cognitive judgments, but are merely indices of individual preferences based upon the relative pleasingness of the various intervals. The significance of this " fact " for any consonance test such as that of Seashore is quite obvious. If many of the judgments secured for the fifty paired-intervals used in this test are not " consonance " judgments (and such is probably the case), then the test is invalid as well as unreliable. A further question perhaps arises concerning the existence of purely cognitive comparison of intervals, apart from the influence of feeling tone. It will be remembered that Guernsey concluded that no such independent judgment was possible. This problem cannot be settled from our data and verbal solutions are meaningless. Further experimentation on carefully selected groups will be necessary before any conclusion can be reached.

[1] A similar view is held by Heinlein. The latter's position is based on the observation of certain types of judgment reversal for consonance judgments incident to reversals in order of presentation. Since these reversals of judgment for certain pairs of intervals which constitute " progressions " and " resolutions " were about what one would expect if consonance judgments are influenced by such factors, Heinlein concluded that they really corresponded to reversals of preference. However, since " preference " judgments were not secured in connection with the consonance judgments, Heinlein's conclusion, although apparently correct, was largely hypothetical.

# CHAPTER VII

## The Influence of Different Criteria on the Consistency of Consonance Judgments

It seems quite likely that the accuracy and the consistency of consonance discrimination are influenced to some extent by the standard or standards used in making the judgments. If the criteria with which the subjects are provided are vague and misleading, the judgments may be expected to be comparatively inaccurate and variable. The criteria most frequently used in judging relative consonance are: *blending, smoothness, fusion,* and *purity.* As Heinlein (**9**, p. 526) has pointed out, these are both relative and indefinite terms, " potent to arouse various types of affective associations in the minds of the listeners." This is particularly true with respect to *fusion.* As was emphasized by Lipps (**16**, pp. 184–209), even Stumpf who held " Verschmelzung " to be synonymous with consonance was somewhat vague as to its exact meaning. In addition to the difficulty of making clear the precise connotations of the above terms, it would seem that the method of employing these criteria also constitutes an important problem to which little attention has been given. Despite the fact that it is often admitted that, severally, these criteria apply with varying degrees of appropriateness to different tonal combinations they are generally referred to in consonance test directions as though they were synonymous. On the contrary, in certain instances they seem to lead to results which are absolutely contradictory to each other. If the Major Third is compared with the Perfect Fourth on the basis of *purity,* the latter interval is generally regarded as the more consonant, whereas if the comparison is made on the basis of *blending,* the third is judged to be the more consonant. The criteria *smoothness* and *fusion* appear to give rise to a similar inconsistency. The Major Third tends more nearly to fuse into a single sound than does the Major Sixth, whereas the latter

interval is decidedly smoother. Such inconsistencies constitute a serious indictment of the usual method of applying criteria. When criteria which, severally, lead to contradictory results are lumped together the subject has no consistent standard by which to make his judgments and under such conditions it is doubtful whether he has any clear idea of the basis upon which his decisions are supposed to be made. Such an attitude could hardly be expected to make for consistency of response.

As early as 1918 Malmberg (17, p. 104) emphasized the importance of the selection of proper criteria with which to provide subjects for making consonance judgments. Prior to that time little attention had been given to the rather obvious complexity of the conditions prevailing in the case of this type of auditory discrimination. Different investigators employed different criteria, with discrepancies in results which might have been expected. These discrepancies manifested themselves in variations in the ranking of the musical intervals within the octave c′c″ with respect to consonance. Authorities agreed that the octave and the fifth could be ranked first and second, respectively, but, for the remaining intervals disagreements occurred. However, Malmberg, noticing that the term " consonance " had been variously defined, undertook an historical review for the purpose of determining what factors had been emphasized by previous investigators. He discovered that consonance had been variously defined in terms of the following criteria: the feeling of satisfaction, agreement of tones, smoothness, fusion, and purity, with slight variants of these. The " most fundamental factor " was found to be *blending*. Malmberg remarks (17, p. 104), " It may never become possible to arrive at absolute agreement in the order of ranking, but it is plain from this brief historical survey that much may be gained in that direction by a clearer conception in regard to the nature of consonance, the analysis of conditions, and specific definition of terms for the purpose of experimental control." Malmberg's most important contribution to experimental technique in the study of comparative judgments of consonance consisted in the definition of three criteria for his subjects, and in the instructions to them to use only one

of the three in any given comparison. The judgments were to be made on the basis of *blending* alone, if the degree of *blending* was perceptibly different for the two intervals compared; if not, *smoothness* was to be employed; and if there was no difference in either *smoothness* or *blending,* the judgment was to be based upon relative *purity.* Malmberg believed that such a procedure would eliminate much of the inconsistency which usually obtained between the standard and the empirical orders of ranking the intervals. Accordingly, a sixty-six unit test (the intervals of the octave c′c″ presented by the method of paired comparison) was presented to the students in an elementary psychology class in the University of Iowa, with instructions to make their judgments according to the above procedure. The results obtained showed a "rather satisfying agreement between the standard order and the empirical order." The few small deviations which did occur were attributed to undue preference for the Major Third in the empirical rankings. However, inasmuch as the testing program was just beginning, the reliability of the consonance test was not directly studied by Malmberg.

The fact that the various criteria used in judging relative consonance do not always lead to consistent results has also been recognized by Seashore. In his 'Psychology of Musical Talent' (26) Seashore adopts *blending, smoothness,* and *purity* as criteria of consonance, *fusion* being omitted "because it does not agree with the ranking in the three criteria here adopted." However, despite this admitted inconsistency, he substitutes *fusion* for *purity* in his Manual of Instructions and Interpretations for Measures of Musical Talent, on the grounds (27) that more comparisons are called for on the basis of *fusion* than on the basis of *purity.* This would seem to indicate that the various criteria should be used selectively, and one wonders why the directions contained in the above manual do not contain such a provision. More recently, however, Seashore has indicated his intention to devise a consonance test one-half of which could be judged on the basis of *smoothness,* and the other half on the basis of *blending.*

The most thorough study of the various criteria in relation to

the accuracy of consonance judgments is that of Guernsey (6). More specifically, this investigation was conducted for the purpose of making an evaluation of fusion, smoothness, and affective-tone as criteria of consonance. Guernsey concluded from her results that pleasantness and unpleasantness are the most legitimate criteria of consonance; tonal fusion was held to be a sensorial rather than a perceptual phenomenon, and smoothness was subject to too great a divergence in connotation in the mind of the listener.

Notwithstanding the importance of the manner of application of criteria for consonance discrimination, a survey of the literature on consonance perception shows that very few investigators have even considered the problem, and that no investigator has made a comprehensive study of it. Malmberg, who instructed his subjects to use a single selected criterion for each comparison in his study of the agreement between the "standard" and the "empirical" rankings of the intervals within the Octave c′ c″, did not study the reliability of empirical judgments. Seashore recognizes the problem, but has nothing to say with respect to its theoretical significance. And Guernsey's investigation is confined chiefly to the study of the relative merits of certain criteria as they pertain to the accuracy of consonance judgments. Hence, in order to determine the comparative effects upon both the accuracy and consistency of consonance discrimination of the use of a single criterion at a time rather than several, two additional types of experiments have been conducted.

### 1. *The "preferential" use of three criteria*

The 50 paired-intervals used in the Seashore Consonance Test were presented to a group of 39 students, by means of the piano. The subjects were instructed to give their decisions on *blending* alone, if the degree of *blending* was perceptibly different; if not, to make their decisions on the basis of *smoothness;* and if there was no difference in either *smoothness* or *blending* to base their decisions on *purity.* The test was first presented on April 17, 1931, in the regular order shown in table 11. A week later the test was again given to the same group under similar conditions.

TABLE 20

*Showing the means, in terms of per cent correct, standard deviations, and reliability coefficients, for the series of paired-intervals in which three criteria were used preferentially by 39 subjects. The total number and the per cent of judgment reversals for the two presentations are also shown.*

| | | | | | Judgment Reversals | |
|---|---|---|---|---|---|---|
| Presentation | Mean | S.D. | $R_{12}$ | P.E. | No. | Pct. |
| 1 | 70.20 | 8.45 | .726 | .05 | 595 | 30.5 |
| 2 | 68.40 | 9.30 | | | | |

The means shown in table 20 for the two applications of the 50 paired-intervals have a final average of 69.30 per cent correct. This average is 5.55 points (in terms of per cent) higher than the average mean for the consonance test shown in table 7. This fact may be taken tentatively (this comparison is made only provisionally since the results shown in tables 7 and 20 are based upon two different groups of subjects) to indicate that subjects tend to be slightly more " accurate " in making consonance judgments when the latter are based on the " preferential " use of *blending, smoothness* and *purity* than when *fusion* and *blending* are used unselectively,[1] *i.e.,* without any directions as to the particular manner in which they are to be applied.

Differences also exist between the " reliabilities " secured by means of the above methods of using criteria. The group which based its decisions on *blending, smoothness,* and *purity* in the preferential manner already described has an *R* (based on gross scores in terms of per cent correct) of .726, whereas the 36 subjects using *fusion* and *blending* unselectively (see table 9) have an average *R* (based on gross scores in terms of per cent correct) of only .52. Furthermore, a comparison of the percentages of judgment reversals for these groups shows, although perhaps less strikingly, the same tendency. The group which based its decisions upon the preferential use of three criteria has 30.5 per cent of judgment reversals (see table 20)

[1] The directions for the " consonance " test which was presented four times to 36 subjects were practically the same as those given by Seashore, with the exception that *smoothness* was not used as a criterion. However, this omission apparently did not affect the results since the latter were found to compare favorably with those secured by other investigators when Seashore's directions were rigidly followed. For this reason the consonance test which was given to the above group has been regarded as comparable to the Seashore test.

for the two applications of the fifty-unit test, as compared with 35.35 per cent for the group[1] which used *fusion* and *blending* unselectively. The foregoing comparisons[2] of means and "reliabilities" indicate that consonance discrimination tends to be more accurate and more consistent when the subjects are instructed to make their decisions on the basis of a single selected criterion than when judging on the basis of two or three criteria. However, not much emphasis can be placed upon these results since the differences are rather small in some instances, and since data from two different groups of subjects are compared.

### 2. *The use of each of two criteria in separate series*

The results secured concerning the "preferential use of three criteria" suggested the possibility that an even better control of the conditions influencing the judgments would be secured if the pairs of intervals could be so grouped that a single criterion—the appropriate one—might be used throughout a given group of paired-intervals. The subject would not have to keep in mind several criteria nor to try to select the one which seemed relevant in a given case. The "Single Criterion" test, already described (*supra,* p. 39), was constructed by classifying the fifty paired-intervals constituting the Seashore Consonance Test into two groups. Part I was composed of twenty-nine pairs of intervals which the experimenter considered could best be judged on the basis of *smoothness.* Part II was composed of twenty-one pairs to be judged on the basis of *blending.* The Single Criterion test was presented twice to a group of 32 subjects. However, as a

[1] In order to make possible certain comparisons in the present chapter, the percentages of judgment reversals for "consonance" tests 1 and 3, and 2 and 4, which were given to 36 subjects (*supra,* p. 30) have been calculated. Tests 1 and 3 were found to have 36.05 per cent of judgment reversals while tests 2 and 4 have 34.66 per cent. The average for these per cents is 35.35.

[2] It has already been pointed out that a reliability coefficient based on scores, in terms of per cent correct, is not a satisfactory index of the consistency of consonance discrimination. It is probable that an index in terms of actual judgment reversals would give a more precise notion of a group's consistency, since in this case actual inconsistencies would not be obscured by the cancellation of errors from one test to another. For this reason the comparison of reliabilities based on the percentage of judgment reversals for each of the above groups is probably more dependable than the one based on the *R*s secured.

means of affording a check upon its comparative reliability, a "Preliminary" test (*supra,* p. 40), similar to that used in the study of the preferential use of three criteria, was also presented twice to the same group.

The results secured for the Preliminary test and the Single Criterion test are shown in tables 21 and 22. According to table 21, the subjects are more "efficient" in the case of the Single Criterion test. A comparison of the means, in terms of per cent correct, for the first presentations of the above tests shows that the mean for the Single Criterion test is 6.18 points higher than that for the Preliminary test. The reliability of this difference, according to the critical index Diff./Sigma Diff. is 3.43. In the case of the second presentations of the above tests, the mean for the Single Criterion test is 3.75 points higher than that of the Preliminary[1] test. Although this difference is not

TABLE 21

*Showing means, in terms of per cent correct, and standard deviations for the two presentations of the "Preliminary" and the "Single Criterion" tests which were given to 32 subjects. The reliabilities of the differences between the means for the respective presentations of these two tests are also shown.*

| | Preliminary Test | | Single Criterion Test | | Comparisons of two types of Tests | |
|---|---|---|---|---|---|---|
| Presentations | Mean | S.D. | Mean | S.D. | Diff. | Diff./Sigma Diff. |
| 1 | 74.50 | 7.70 | 80.68 | 6.74 | 6.18 | 3.43 |
| 2 | 74.00 | 7.80 | 77.75 | 7.54 | 3.75 | 1.97 |

TABLE 22

*Showing the reliabilities for the two presentations of the "Preliminary" and the "Single Criterion" tests given to 32 subjects.*

| | | | Judgment Reversals | |
|---|---|---|---|---|
| Tests | *R* | P.E. | No. | Pct. |
| Preliminary Test | .439 | .09 | 378 | 23.62 |
| Single Criterion Test | .580 | .07 | 328 | 20.50 |

[1] It will be recalled that the Preliminary test differed from the Seashore Consonance test chiefly in point of directions. In the case of the Preliminary test, the subjects were directed to use *smoothness* and *blending* in the preferential manner already described (*supra,* p. 38). In the case of the Single Criterion test, the pairs of intervals were so arranged that the subjects could apply a single criterion at a time. They were provided with this criterion by the experimenter. Both the Preliminary test and the Single Criterion test were played on the piano.

statistically reliable (Diff./Sigma Diff. is 1.97), it is sufficiently large to make it highly probable that the Single Criterion test presents a situation which is more conducive to accurate discrimination than that provided by the Preliminary test. These results indicate that when the paired-intervals are so arranged that the subject has only to apply a single criterion at a time the psychological "set" is more conducive to "accurate" consonance discrimination than when he is required to select the appropriate criterion for each pair before making the comparison. When we take into consideration the short time usually allowed for each consonance judgment it is not surprising to find that subjects are more "accurate" in their judgments when they have only to apply a single criterion than when they are required within a few seconds of time to determine which of several criteria is the appropriate one and also to apply it.

The comparative reliabilities of the Preliminary and of the Single Criterion tests are shown in table 22. These have been computed on two bases: first, the reliability coefficients have been calculated by correlating the scores, in terms of per cent correct, of the 32 subjects made on the two presentations of each test: and second, the per cent of judgment reversals for each test has also been computed.

It will be observed at the outset that the $R$ secured for the Preliminary test (.439), which called for the preferential use of *smoothness* and *blending,* is even lower than the average $R$ obtained for the Seashore test (.52, see table 9) in which two criteria were used unselectively. This apparently contradicts the conclusion drawn from the previous experiment that subjects are more consistent in judging relative consonance when using criteria preferentially than when using them unselectively. However, this contradiction is only an apparent one. It has been pointed out repeatedly that a reliability coefficient based on the correlation of scores, in terms of per cent correct, is not a reliable index of subjects' consistency in making consonance judgments. The present case furnishes further evidence in support of this statement. A comparison of the average per cent of judgment reversals for the four applications of the Seashore test (35.35)

with that secured for the Preliminary test (23.62) shows that the subjects were actually more consistent in the case of the latter test. Since these figures are based on two groups of subjects the disparity between them should not be too greatly emphasized. However, they do serve to indicate that when the comparisons are made on the basis of actual judgment reversals the results obtained for the present experiment do not contradict our previous findings. In this connection it is important for the reader to note that although both the reliability coefficients and the per cents of judgment reversals are usually shown for the various consonance series, our conclusions are based almost invariably upon the latter indices.

Examination of table 22 shows that regardless of whether the reliabilities of the Single Criterion test and the Preliminary test are stated in terms of reliability coefficients or in terms of per cent of judgment reversals, the former test is the more reliable of the two. This test has an *R* of .58 as compared with an *R* of .439 for the Preliminary test. The reliability of the difference between these coefficients, according to the critical index Diff./Sigma Diff., is 11.75. Although less disparity is shown between the " reliabilities " of the above tests when the comparisons are made in terms of per cent of judgment reversals, the tendency manifested is the same as that just noted when the *R*s for these tests were compared. The Single Criterion test has 20.50 per cent of judgment reversals as compared with 23.62 per cent for the Preliminary test. Thus, either of the above methods of comparison shows the Single Criterion test to be the more reliable. This means that subjects are more consistent in judging relative degrees of consonance when the pairs of intervals are so arranged that the subjects have only to apply a single criterion at a time than when they are required to select the most appropriate criterion for each comparison before making their judgments.

The foregoing experiments show that consonance discrimination is conditioned by the criteria upon which comparisons of relative consonance are based. Regardless of the groups involved, each time the technique of applying criteria was refined with the view to reducing the complexity of the act of judging,

more " accurate " and consistent results were obtained. The first " consonance " series (consisting of four applications of Seashore's fifty paired-intervals) in which two criteria were used unselectively had an average mean of 63.75 per cent correct (average of the means shown in table 7), and an average of 35.35 per cent of judgment reversals. The second group which was instructed to apply three criteria preferentially had an average mean of 69.30 per cent (based on the means shown in table 20), and 30.5 per cent of judgment reversals. The third group of subjects, using two criteria preferentially, had an average mean of 74.25 per cent (based on the means shown in table 21), and 23.62 per cent of judgment reversals. When the pairs of intervals were so arranged that the third group had only to apply a single criterion at a time the subjects had an average mean of 79.21 per cent correct, and only 20.50 per cent of judgment reversals. Although these results were secured from different groups of subjects they point to the same general conclusion, namely, that consonance discrimination is influenced by the manner in which criteria are employed in making comparisons of relative consonance.

Inasmuch as the above data were obtained from three groups of subjects, generalizations cannot be too freely made. However, there are good reasons for holding the above differences in accuracy and consistency to be due to the different methods employed in using criteria rather than to the mere fact that the data were secured from different groups. The difference in means between the group of subjects using two criteria unselectively and that using a single criterion at a time is 15.46 points, in terms of per cent correct; the corresponding difference between the per cents of judgment reversals is 14.85 points. Now it seems improbable that these rather large differences are due merely to the fact that the data secured were for separate groups, since all three samplings consisted of unselected subjects from classes in general psychology. Furthermore, the average mean for the group using a single criterion at a time is considerably higher than that reported by any investigator for unselected groups, and slightly higher than those reported by Larson for

35 subjects selected on the basis of musical training (*supra,* p. 45). And again, the results shown for the Preliminary and the Single Criterion tests were based on the same group of subjects. In addition to the foregoing reasons, a consideration of the conditions obtaining for the first set of " consonance " tests in which two criteria were used unselectively shows that the comparative " inaccuracy " and unreliability of the results secured for these tests are attributable in part to the failure to secure proper control, with respect to the use of criteria. The stimuli used in the case of the first " consonance " tests consisted of fifty paired-intervals played on a phonograph. Each interval was sustained approximately two seconds, one second being allowed between the intervals of a pair, and two seconds between each pair. During this brief sustension of an interval the subject was expected to disregard the most obvious characteristic of any tonal combination (affective-tone), and to make a cognitive judgment on the basis of several variously appropriate criteria. Under such conditions it is doubtful whether the average subject has in mind any particular standard for judgment throughout an entire series, much the less from one series to another. At least, the attempt to judge relative consonance under such conditions resulted in an average mean of only 63.75 per cent correct, and in the making of 35.35 per cent of judgment reversals. This low mean and high percentage of judgment reversals indicate that in many instances the judgments were largely the result of chance factors. The foregoing facts make it fairly obvious that any test such as the Seashore test which fails to provide subjects with a satisfactory basis for judging relative consonance—*i.e.,* which fails to control an important set of conditions—will necessarily yield " inaccurate " and inconsistent results.

## CHAPTER VIII

### General Summary

The main purpose of this study has been to clarify somewhat the status of the problem of consonance discrimination—both with respect to the numerous theoretical "explanations" which have been proposed and as regards experimental fact. No unanimity of opinion existed with regard to the definition of consonance, and writers were even unable to agree as to the facts for which any theory of consonance perception must account. The majority of the theories dealing with this phenomenon were found to be based mainly upon *a priori* speculation, and consisted of profitless verbalisms, over-simplifications, and partial explanations, each emphasizing certain facts which were often regarded as unimportant by the authors of other theories. In fact no "explanation" of consonance was found to be entirely adequate. As regards the experimental aspect of the present problem, it was found that investigators had been unable to secure consistent judgments of paired tonal stimuli, and that furthermore, results of the various studies were often contradictory. As previously pointed out, this inability to secure consistent judgments of relative consonance indicates a lack of proper experimental control of the conditions influencing consonance discrimination and therefore constitutes a general breakdown of scientific method in the investigation of an important problem of perception. Thus the problem concerned with the control of conditions influencing consonance discrimination was held to take precedence over all the other problems connected with this phenomenon, since the study of almost any aspect of consonance perception is dependent upon the ability of the investigator to obtain consistent judgments of relative consonance.

Recognizing the importance of securing consistent judgments of relative consonance the present investigation undertook the study of three sets of conditions of consonance discrimination:

(1) the difficulty of the comparisons called for; (2) affective-tone; (3) the criterion or criteria used. It was not expected in this initial study that we could determine the influence of these three sets of conditions upon consonance discrimination and at the same time effect such control over them as to avoid any inconsistency. Rather it was hoped to discover what effect, if any, these sets of conditions might have upon consonance discrimination and to indicate the general manner in which they might be controlled, so as to reduce materially the inconsistency which has regularly attached to such judgments. Fortunately, however, some progress has been made in the securing of more consistent judgments. In studying the foregoing sets of conditions it has been necessary to devise certain methods which should be of value in a further, more intensive study of the present problem. Furthermore, the attempt has been made to indicate just wherein certain of the traditional means of attacking the problem of consonance perception (*e.g.*, the determination of the consistency of consonance judgments by correlating gross scores in terms of per cent correct for two applications of a series of paired-intervals) have proven inadequate, and also to show how the application of such methods has been responsible for much of the confusion which obtains with respect to the reliability of consonance judgments. In short, the present investigation proposed to study the effects of three important sets of conditions upon consonance discrimination and at the same time to develop certain methods which might be used with profit in further work on this problem.

The results secured in the present investigation indicate that consonance discrimination is affected to some extent by the difficulty of the comparisons made, the consistency of the subjects' judgments tending to vary inversely with the difficulty of the comparisons. The greatest inconsistency was found to occur in connection with those combinations which are so difficult as to make the subjects' judgments largely matters of "chance." However, certain "very difficult" pairs of intervals showed two opposite trends: (1) the judgments seemed to be conditioned by an attitude of guessing on the part of the subjects, which nat-

urally made for unreliability, and (2) certain constant errors occurred and so made for a lower percentage of judgment reversals. With respect to the latter tendency it was pointed out that probably certain factors (*e.g.*, affective-tone) so bias the subject that he is unable to make a detached judgment on the basis of relative consonance. Thus it is seen that such an apparently homogeneous condition as "difficulty of comparisons" is quite complex, embodying a variety of factors. As previously indicated, the failure to take this fact into account is doubtless partly responsible for the low reliability of the Seashore test. Approximately 50 per cent of the paired-intervals contained in the latter test were found to present such difficulty that the judgments secured for them seem to be largely the result of chance factors. The present investigation has demonstrated that such "chance judgments" are very unreliable.

It was shown in our historical introduction that while writers on the subject are fairly well agreed as to the distinction between consonance and pleasingness the probable influence of affective-tone upon consonance discrimination has been the subject of considerable controversy. Heinlein holds that so-called consonance judgments are conditioned by feeling-tone, whereas Larson maintains that under conditions such as are prescribed in the Seashore Consonance Test they are not affected to any appreciable degree by this element. However, it was pointed out that the claims of both of these investigators were based upon data secured from subjects under a single set of conditions rather than upon judgments obtained under two sets of conditions, *i.e.*, under directions calling for judgments on the basis of relative pleasingness during one application of a series of paired-intervals, and on relative consonance during another. Thus the contention of each of these writers was discovered to rest largely upon speculation as to how the subjects *would* react in case their judgments were influenced by affective-tone. In the present study comparative judgments were secured under *both* sets of conditions, for the fifty paired-intervals constituting the Seashore test. The results obtained are shown in tables 11 and 19, and are fairly decisive with respect to the question in hand. Our analysis of these

tables showed that subjects' judgments tend to be "correct" when the more pleasing interval is also the more consonant one, but "incorrect" when the less consonant of two intervals is decidedly the more pleasing of the two, thus indicating that feeling-tone is the real basis of many of the decisions made. However, this does not mean that it is impossible for subjects to disregard affective-tone, and make cognitive judgments of relative consonance, since our results showed that in certain instances some of the subjects were able to do so. It does mean, however, that consonance discrimination which calls for a purely cognitive judgment is, despite certain precautions to the contrary, often influenced by the affective quality of the intervals presented for comparison. Before consonance perception can be extensively investigated some means must be devised for controlling or correcting for the influence of this affective element. It is possible that the probable influence of affective-tone upon the judgments of any paired-interval might be determined by comparing a large number of judgments secured under the two sets of conditions just mentioned. If, for example, "consonance" judgments obtained by comparing the Perfect Fourth with the Major Third were found to be 25 per cent more "accurate" when the more consonant of the two intervals is also the more pleasing, then we would have an approximate measure of the influence of affective-tone upon the judgment of these two intervals. However, this problem can be settled only by further investigation.

The act of judging relative consonance has been further complicated by the failure to provide subjects with a suitable basis upon which to make their judgments. The criteria *fusion, blending, smoothness,* and *purity,* which have been employed regularly as consonance criteria, have non-auditory connotations, hence their precise meanings in relation to tonal unity are probably not clear to the average subject. However, it has been shown that this confusion is increased when one is required (as is usually the case) to regard these criteria as synonymous, although they severally apply with varying degrees of appropriateness to various tonal combinations. In certain instances the

use of two of the criteria leads to contradictory judgments. For example, the writer has observed that when the Major Third is compared with the Perfect Fourth, on the basis of *purity,* the latter interval tends to be considered the more consonant of the two, whereas, when the comparison is made on the basis of *blending,* the Major Third is judged to be the more consonant. That much of the inaccuracy and inconsistency frequently observed in results secured for consonance series is due to confusion occurring in the act of applying criteria has been rather clearly shown by the present study. When certain of the above criteria are used selectively, more " accurate " and consistent results are obtained than when the various criteria are used collectively and indiscriminately. And still more satisfactory results are secured when the paired-intervals are so arranged that the subject has only to apply a single appropriate criterion (which is provided by the experimenter) at a time. These facts indicate plainly the dependence of consonance discrimination upon the " set " of the subject, as regards his technique in applying criteria. The failure to secure in the subject a definite and unambiguous attitude, in terms of the " criteria " to be employed, constitutes probably one of the chief defects of most previous experimental studies of consonance.

In general the present study has shown that " consonance " is a complex perceptual phenomenon which cannot be adequately accounted for or profitably investigated on the basis of any of the traditional, over-simplified hypotheses. Both the theoretical and the experimental studies which have been based upon such assumptions have been of questionable scientific value. The former have been, for the most part, ideal constructions bearing little relation to the actual perceptual process, while those experimental studies which have assumed that consonance discrimination involves merely the direct response of the sensory mechanisms to stimuli which could be compared in terms of a definite " linear " differential have contributed little or nothing to our understanding of the phenomenon. While it is unlikely that we are in possession of enough facts to justify any sweeping generalization as to the explanation of consonance, it is apparent

that future studies, both theoretical and experimental, must take into consideration the complexity of conditions by which consonance perception is influenced. Among the experimental problems which emerge from the present study the following would seem to be outstanding:

(1) Further intensive analysis of the operations of "cognitive" versus "affective" attitudes in determining judgments of relative consonance of specific paired-interval combinations.
(2) The influence of affective-tone upon "consonance" judgments of musically trained and untrained subjects, respectively.
(3) The effects of practice (including specific instruction) upon success in maintaining a "cognitive" attitude in the case of judging combinations shown here to be, in general, directly influenced by "resolution."
(4) The effects of systematic alterations of the order of presentation of intervals in a series, to check on the effects of "progression."
(5) Possible variations in "consistency" of judgments of "difficult" combinations with alterations in the technique of presentation which would allow such a combination to be sounded several times, if desired, thus relieving the subject of the impulsion to make a hurried judgment such as the phonograph method requires.
(6) The effects of differences in the *timbre* of the tones upon the "difficulty" of comparing certain intervals, with special attention to differences in "roughness" due to variations in beats among overtones and difference tones.
(7) Careful study of the extent to which the criterion or criteria provided operate *explicitly* to determine judgments in regard to specific combinations.
(8) An empirical determination of the appropriations of each of the several criteria of consonance as a standard for judging specific intervals. This would involve a study of carefully instructed subjects with reference to agreement and consistency in the use of such criteria as *blending, smoothness, fusion,* and *purity,* respectively, for judgments of combinations for which each seemed especially appropriate.

# CHAPTER IX

## Conclusions

The present investigation is to be regarded as a preliminary, analytical study of some of the factors influencing consonance discrimination. It is fully realized that an investigation such as the present one is not exempt from the dangers due to the variability of this type of response simply because it is devoted to the study of that variation. With these qualifications in mind the following conclusions are drawn on the basis of the foregoing study:

1. Subjects tend to make lower " scores " upon the repetition of both " consonance " and " preference " series. Our 36 subjects from whom data were secured for four applications each of the " consonance " and of the " preference " tests made scores which averaged about 5 per cent less (in terms of per cent correct) on the fourth application of each type of test than those made on the initial applications. This was perhaps to have been expected, since many of such comparisons are difficult and require the maintenance of an alert, highly discriminative attitude. This is difficult to secure, since it depends upon unusual motivation and perhaps upon a type of training not possessed by the average subject.

2. Musically trained subjects are slightly more " accurate " and consistent in judging relative consonance than untrained subjects. Our musically trained and untrained groups had respective averages of 65.2 and 62.5 per cent correct for the " consonance " judgments. The corresponding per cents in the case of the " preference " judgments were 68.5 and 63.7. The average *R*s secured for these respective groups were .543 and .466 for the " consonance " series as compared with .625 and .525 for the " preference " series. These results agree in general with those secured by other investigators.

3. A subject's " score " obtained for the Seashore Consonance Test cannot be considered as a satisfactory index of his consonance discriminability. The notion that it is possible, within a few minutes, to measure a complex capacity which seems to vary with so many conditions and which is not a " sense " in the true meaning of the term has little factual support. Not only are the diversely conditioned judgments secured on the Seashore test and generalized into such a " score " too inconsistent to be of practical predictive value, but in many instances they are determined chiefly on the basis of the relative pleasingness of the intervals presented for comparison, and hence are not cognitive judgments.

4. The number of judgment reversals made by a group for a series of paired-intervals is probably a more accurate index of consistency than is a reliability coefficient based on " scores " in terms of per cent correct. The number of judgment reversals increases directly with each inconsistent response, whereas the " per cent correct " score permits compensation of a reversal scored as an error by another reversal scored as " correct."

5. Consonance discrimination is not the absolutely irregular, fortuitous phenomenon that the usual type of " reliability coefficients," based on gross scores, would seem to indicate. When the reliability coefficients were based on the error-frequency per combination for our 36 subjects the " consonance " judgments had an average $R$ of .80, and the " preference " judgments had an average $R$ of .90. These results indicate a very high *group consistency* in regard to the relative difficulty of the paired-interval comparisons.

6. Table 11 of the present study contains data of a nature not hitherto available in the literature, which permit several types of analysis made for the first time in this study. The numbers and percentages of errors for eight applications of each combination in the Seashore test are shown in this table; the " error-frequencies " for both " consonance " and " preference " tests are given. In addition to showing the relative difficulty of each of the fifty paired-intervals constituting the Seashore test, the above table has made possible the comparison of certain types of judgment

reversals incident to changes in order of presentation of intervals within a pair, for both "consonance" and "preference" judgments. In showing that subjects' judgments are, in general, considerably more "accurate" when pleasantness favors the "correct" judgment than when it favors the "incorrect" one the present study has presented decisive evidence bearing upon the question of the influence of affective-tone upon consonance discrimination.

7. The consistency of consonance discrimination tends to vary inversely with the difficulty of the comparisons called for. The group of 36 subjects taking the Seashore Consonance Test was found to have only 19.4 per cent of judgment reversals for the four applications of the ten easiest combinations contained in this test, whereas 37.3 per cent of judgment reversals were made for the ten most difficult comparisons. In some instances, however, certain very "difficult" comparisons seem to engender some degree of consistency since they make for constant errors. It should be noted here that the phrase "paired-interval difficulty" is equivocal. It may mean actual confusion in comparing intervals which are very similar in consonance value, or it may refer to the inability of the subject to disregard irrelevant factors which make his judgment "easy" (in the sense that it is made with confidence), yet "erroneous." "Paired-interval difficulty" is referred to here in terms of the per cent of erroneous judgments made for the combinations, irrespective of cause.

8. Although "preference" judgments tend to be more "accurate" and more consistent (in terms of scores and reliability coefficients) than so-called "consonance" judgments, the differences are so slight as to suggest that in many instances the "sets" of the subjects are about the same for these supposedly different types of response. This conclusion is supported by analyses made of the effect of order of presentation of intervals upon consonance and preference judgments which show that the frequency with which an interval is regarded as the more consonant one of a pair tends to vary directly with the frequency with which it is regarded as the more pleasing interval.

9. The "accuracy" and consistency of consonance judgments are conditioned to some extent by the standards (criteria) used

in making the comparisons. The group of subjects which used two criteria unselectively had an average mean of 63.75 per cent correct, and an average of 35.35 per cent of judgment reversals. The group using three criteria preferentially had an average mean of 69.30 per cent correct, and 30.5 per cent of judgment reversals. The third group of subjects, using two criteria preferentially, had an average mean of 74.25 per cent, and 23.62 per cent of judgment reversals. When the pairs of intervals were so arranged that the third group had only to apply a single criterion at a time the subjects had an average mean of 79.21 per cent correct, and only 20.50 per cent of judgment reversals. These data might be said to constitute a 'consilience of results' in that they point in the same general direction. Each time the act of applying criteria was further simplified, more "accurate" and consistent results were secured, thus indicating that consonance discrimination is conditioned by the criteria used in judging relative consonance. In general, the more definite and specific the "set" of the subject the more efficiently is he able to discriminate relative consonance.

10. Although any sweeping general theory of consonance perception is perhaps premature, the present study has shown that comparative judgments of consonance are complexly conditioned phenomena—too complexly conditioned to be accounted for by any theory which regards "consonance" as a simple, all-or-none sensory process. With respect to further experimental work, several important problems have been suggested (*supra,* p. 95); intensive analytical study on each of these and of other similar, restricted problems, should be the aim of future investigators in this field.

## References

1. Brown, Euleen. Consonance preferences among whites and negroes. M.A. thesis, George Peabody College, 1927, 1–67.
2. Emerson, L. E. The feeling value of unmusical intervals. *Harvard Psychol. Studies,* 2, 269–270.
3. Farnsworth, P. R. An historical, critical, and experimental study of the Seashore-Kwalwasser test battery. *Genet. Psychol. Monog.,* 1931, 9, No. 5, 297–389.
4. Gaw, E. A. A revision of the consonance tests. *Psychol. Monog.,* 1918, **25**, No. 108, 134–147.
5. Grove, George. *Dictionary of music and musicians.* New York: The Macmillan Co., 1928, **4**, p. 366.

6. GUERNSEY, M. The rôle of consonance and dissonance in music. *Amer. Jour. Psychol.*, 1928, **40**, 173–204.
7. GUTHRIE, E. R., and MORRILL, H. The fusion of non-musical intervals. *Amer. Jour. Psychol.*, 1928, **40**, 624–625.
8. HEINLEIN, C. P. An experimental study of the Seashore consonance test. *Jour. Exper. Psychol.*, 1925, **8**, 408–433.
9. HEINLEIN, C. P. Critique of the Seashore consonance test: a reply to Dr. Larson. *Psychol. Rev.*, 1929, **36**, 524–542.
10. HELMHOLTZ, H. *Sensations of tone.* New York: Longmans, Green and Co., 1912 (translated from 4th German edition by A. J. Ellis), pp. xix + 556.
11. HUNTER, WALTER S. The psychological study of behavior. *Psychol. Rev.*, 1932, **39**, 1–24.
12. KRUEGER, FELIX. Consonance and dissonance. *Jour. Phil., Psychol., and Sc. Method,* 1913, **10**, 158–160.
13. LANIER, L. H. Prediction of the reliability of mental tests and tests of special abilities. *Jour. Exper. Psychol.*, 1927, **10**, 69–113.
14. LARSON, D. L. An experimental critique of the Seashore consonance test. *Psychol. Monog.*, 1928, **38**, No. 176, 49–81.
15. LEIBNIZ, G. W. Principes de la nature et de la grâce, Leibnitii epistolae ad diversos. Leipsic: Breitkopf, pp. 240 (Epistel 154).
16. LIPPS, THEODOR. *Psychological studies* (translated by H. C. Sanborn). Vol. II of Psychology Classics. Baltimore: Williams and Wilkins, 1926, pp. 138–265.
17. MALMBERG, C. F. The perception of consonance and dissonance. *Psychol. Monog.*, 1918, **25**, No. 108, 93–133.
18. MOORE, H. T. The genetic aspect of consonance and dissonance. *Psychol. Monog.*, 1914, **17**, No. 73, 1–68.
19. OGDEN, R. M. *Hearing.* New York: Harcourt, Brace and Co., 1924, pp. xiii + 351.
20. PETERSON, J. A functional view of consonance. *Psychol. Rev.*, 1925, **32**, 17–33.
21. POLE, W. P. *The philosophy of music.* New York: Harcourt, Brace and Co., 1924, pp. xiii + 334.
22. PRATT, C. C. Quarter-tone music. *Jour. Genet. Psychol.*, 1928, **35**, 286–292.
23. RUCH, G. M., and STODDARD, G. D. *Tests and measurements in high school instruction.* Yonkers, N. Y.: World Book Co., 1927, pp. xvi + 373.
24. SCHOPENHAUER, A. *Die Welt als Wille und Vorstellung.* Zweite Auflage, Dritter Band. Leipzig: Brockhaus, 1891, pp. 511–523.
25. SEASHORE, C. E. Measurement of pitch discrimination: a preliminary report. *Psychol. Monog.*, 1910–1911, **13**, No. 53, 21–63.
26. SEASHORE, C. E. *The psychology of musical talent.* Boston: Silver, Burdett and Co., 1919, pp. xvi+288.
27. SEASHORE, C. E. Measures of musical talent: A reply to Dr. C. P. Heinlein. *Psychol. Rev.*, 1930, **37**, 178–182.
28. STANTON, H. M. Seashore measures of musical talent. *Psychol. Monog.*, 1928, **39**, No. 178, 135–144.
29. STUMPF, KARL. *Tonpsychologie,* Erster Band. Leipzig: Verlag von S. Hirzel, 1883, pp. xiv + 563.
30. WEAVER, A. T. Experimental studies in vocal expression. *Jour. Appl. Psychol.*, 1924, **8**, 23–51.
31. WUNDT, W. *Grundzüge der physiologischen Psychologie.* Fünfte völlig umgearbeitete Auflage. Zweiter Band. Leipzig: Verlag von Wilhelm Engelmann, 1902, pp. viii + 686.